MECHANISMS AND FUNCTIONAL ROLES OF NUCLEAR RESPIRATORY FACTOR 1 (NRF1) BINDING SITES IN THE HUMAN GENOME

by

Wan Zhu

B.S., Biosciences, Sun Yat-sen University, China, 2005

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This dissertation was presented

by

Wan Zhu

It was defended on

July 14th, 2011

and approved by

Robert D. Nicholls, D. Phil., Professor, Department of Medical Genetics, Children's Hospital of Pittsburgh, UPMC and Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

Robert E. Ferrell, Ph.D., Professor, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

Eleanor Feingold, Ph.D., Professor, Department of Human Genetics and Biostatistics, Graduate School of Public Health, University of Pittsburgh

Susanne M. Gollin, Ph.D., Professor, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

MECHANISMS AND FUNCTIONAL ROLES OF NUCLEAR RESPIRATORY FACTOR 1 (NRF1) BINDING SITES IN THE HUMAN GENOME

Wan Zhu, Ph.D.

University of Pittsburgh, 2011

Genome-wide studies have suggested that NRF1 regulates transcription of ~5-6% of human genes, including nuclear genes encoding mitochondrial products. My thesis focus is in neural systems in which NRF1 is a master regulator.

Prader-Willi syndrome (PWS) results from genetic loss of function of an imprinted domain in human chromosome 15q11.2. I confirmed NRF1 regulation of ~83% of PWS-region genes using chromatin immunoprecipitation (ChIP). Further studies focused on evolution of this region. Uniquely in marsupials, *SNRPN* and the ancestral *SNRPB'* gene are adjacent each with an intronic snoRNA paralog. Based on molecular phylogenetics, a model is proposed for origin of each PWS snoRNA from a single ancestral snoRNA. Thus, most extant eutherian PWS genes originated by stepwise duplication and divergence over the past ~180 million years.

Circadian rhythms regulate organismal physiology in a 24 hour day-night cycle. Functional NRF1 binding sites in promoters/enhancers were found for ~56% of circadian regulatory genes using bioinformatics, ChIP, *NRF1* siRNA assays, and luciferase reporter constructs having significantly reduced transcriptional activity on mutation of NRF1 sites. Further, co-immunoprecipitation showed that NRF1 and the phosphorylated, active form of CLOCK interact in a molecular complex. In serum-induced NIH3T3 cells with circadian oscillations of *Dbp* and *Nr1d1* mRNA, *Nrf1* mRNA and protein levels show ultradian

oscillations. Hence, NRF1 regulates numerous circadian regulatory genes and interacts with CLOCK, suggesting multiple roles in circadian biology.

Additional studies included finding that NRF1 regulates ~45% known hereditary spastic paraplegia (HSP) genes, that NRF1 activates its own transcription, and that the number of NRF1 sites determine the degree of transcriptional activation.

In summary, NRF1 is a master regulator in PWS, circadian rhythms, and HSP. Identification of NRF1 target genes and mechanisms will lead to an understanding of the evolution, functions, disease processes, and therapeutic targets within gene regulatory networks involving NRF1. Circadian rhythms are disrupted by travel, shift-work, and in illness, including infection, psychiatric and sleep disorders, obesity, diabetes, and cancer. Consequently, understanding body clocks will provide insights into the pathogenesis of these disorders and potentially lead to improved treatment and prevention options, which will have enormous public health impact.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Foremost, I would like to show my earnest appreciation to my mentor, Dr. Robert D. Nicholls, for his continuous encouragement in my research work, his constructive guidance in writing and his intelligence that impresses me in so many ways. I would like to express my sincere gratitude to Dr. Eleanor Feingold, Dr. Robert Ferrell, Dr. Susanne Gollin and Dr. Panayiotis (Takis) Benos, who serve as my committee members and give comments to my thesis.

Next, I would like to acknowledge my current and former co-workers in Nicholls' laboratory, including Dr. Mihaela Stefan, Dr. Brian J. Henson, Mr. Shanping Li and Ms. Jaime L. Rupert, for their support and work to complete my thesis. I would also like to thank Dr. Paul Samollow, Dr. Panayiotis V. Benos, Ms. Wei Wang, Dr. Yudong Wang and Dr. Marie Semele, for their advices in my research work.

Special thanks to Dr. Eleanor Feingold and Dr. Candace M. Kammerer for their advice and support when I had conflicts between my study and my life.

Last but not least, I would like to express the deepest love to my parents and my husband. Without their support and encouragement, I would not be able to accomplish this work. I want to dedicate my thesis to my dearest daughter, Annabelle, who brings so much happiness and new meaning to my life!

Abbreviations:

AS	Angelman syndrome
ChIP	Chromatin immunoprecipitation
ChIP-chip	Chromatin immunoprecipitation on chip
CREB1	cAMP responsive element binding protein 1
C_T	Threshold cycle
DMR	Differentially methylated region
ER	Endoplasmic reticulum
ESRRA	estrogen-related receptor alpha, also known as ERRα
GABP	GA binding protein transcription factor, alpha subunit
GRN	Gene regulatory network
HSP	Hereditary spastic paraplegias
IC	Imprinting center
lincRNA	Long intergenic noncoding RNA
miRNA	microRNA
NRF1	Nuclear respiratory factor 1
PPARs	Peroxisome proliferator-activated receptors
PWS	Prader-Willi syndrome
PyPu	Pyrimidine and purine
QRT-PCR	Quantitative real-time PCR
R	Adenine or guanine
RCR	Regional control region

snoRNAs	Small nucleolar RNAs
TF	Transcription factor
TFAM	Mitochondrial transcription factor A
TSS	Transcription start site
WT	wildtype
YY1	Yin Yang 1
Y	Cytosine or thymine (uracil)
ΔG	Gibbs energy (thermodynamics)

1.0 INTRODUCTION

1.1 NUCLEAR RESPIRATORY FACTOR 1

1.1.1 Gene and protein

Nuclear respiratory factor 1 (NRF1, also originally named α-Pal) is a highly conserved, phosphorylated nuclear protein (**Figure 1.1**). It is encoded by a gene with 11 exons including a non-coding first exon. NRF1 was first found to regulate genes in mitochondrial respiratory function (Scarpulla, 2008). Subsequent research has found that NRF1 recognition sites occur in many nuclear genes encoding factors involved in mitochondrial function and biogenesis (Evans & Scarpullar, 1989; 1990; Virbasius *et al*., 1993; Scarpulla, 2008; Kelly & Scarpulla, 2004), cell cycle regulation (Cam *et al*., 2004; Efiok & Safer, 2000), and neuronal functions (Smith *et al*., 2004; Chang *et al*., 2005). As a transcription factor (TF), NRF1 forms homodimers and recognizes a unique DNA binding sequence, YGCGCAYGCGCR, (Scarpulla, 2006). This motif consists of a palindromic, direct repeat, and $(PyPu)_n$ sequence structure, but the relative contributions of each of these alternative components to binding affinity are not known. In virtually all described circumstances, NRF1 functions as a positive TF to upregulate gene expression (Scarpulla, 2008), although binding of chicken NRF1 at the transcription start site (TSS) of the histone H5 gene inhibits transcription (Gómez-Cuadrado *et al*., 1995), as may regulation by the sea urchin NRF1

(P3A2) at the *CyIIIa* gene (Bogarad *et al*., 1998). The regulatory roles of NRF1 may thus be complex and depend on the context of the regulatory elements.

NRF1 has an ancient origin in animals, with orthologs identified in mouse (Nrf1), chicken (NRF1) (Gómez-Cuadrado *et al*., 1995), zebrafish (nrf1) (Becker *et al*., 1998), sea urchin (P3A2) (Calzone *et al*., 1991; Höög *et al*., 1991), and *Drosophila* (EWG) (DeSimone & White, 1993), all that share a highly conserved DNA binding region (**Figure 1.1**). Recent molecular phylogenetic studies have identified NRF1 orthologs in all animals with whole genome sequences except for the worm *C. elegans* (R.D. Nicholls unpublished data). The DNA binding domain of NRF1 is unique and not shared with any other TF class. The constrained amino acid sequence of the DNA binding domain of NRF1 suggests that a similar binding site is recognized across the animal kingdom. Indeed, *Drosophila* EWG binds to an identical DNA motif as mammalian NRF1 (Fazio *et al*., 2001).

Knock-out and mutation studies in animals indicate that NRF1 has irreplaceable functions in animal development and differentiation. In mouse, a homozygous *Nrf1* knockout mutation is lethal between embryonic days E3.5 and E6.5 (Huo & Scarpulla, 2001). Although *Nrf1* -/- blastocysts appear normal in morphology, these could not grow *in vitro* in cell culture and had significantly reduced amounts of mitochondrial DNA as well as decreased mitochondrial staining with a fluorescent dye that measures mitochondrial membrane potential (Huo & Scarpulla, 2001). Disruption of zebrafish *nrf1* is lethal at the larval stage and also shows a specific retinal phenotype with loss of all photoreceptor neurons and their precursors (Becker *et al*., 1998). Similarly, loss of the sea urchin P3A2 is embryonic lethal (Bogarad *et al*., 1998). Partial loss-of-function mutations of *EWG* in fly show severe neuronal and muscle defects while complete loss-of-function mutations are embryonic lethal (Fleming *et al*., 1989; DeSimone & White, 1993; Haussmann *et al*., 2008).

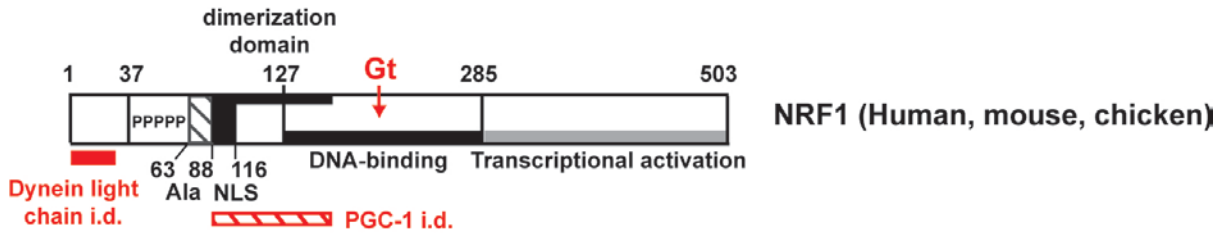


Figure 1.1. Structure of the NRF1 polypeptide. NRF1 from human, mouse, and chicken are 99-100% identical over the entire sequence, compared with the orthologs (not shown) from sea urchin (P3A2; 67.5% identity in the DNA binding domain) and *Drosophila* (EWG; 59.7% identity in the DNA binding domain). The position of polypeptide structural domains are indicated, including alanine (Ala)-rich (hatched box), nuclear localization signal (NLS, black vertical box), DNA binding (black horizontal box), dimerization (small black horizontal bar) and transcriptional activation (grey box) domains. NRF1 has 5 serine residues that are phosphorylated (P), and four of these serines are conserved in P3A2 and EWG. Other symbols are: interaction domains (i.d.) for dynein light chain and PPARGC1A/PRC; Gt: position of the RRF126 gene-trap β-geo fusion truncating the mouse Nrf1 polypeptide. Modified from: Höög *et al*., 1991; Efiok *et al*., 1994; Gómez-Cuadrado *et al*., 1995; Gugneja *et al*., 1996; Gugneja *et al*., 1997; Herzig *et al*., 2000; Andersson & Scarpulla, 2001; Huo & Scarpulla, 2001; Nicholls RD, unpublished data.

1.1.2 NRF1 and mitochondria

The mitochondrion is a membrane-enclosed organelle found in most eukaryotic cells. Known as "cellular power plants", mitochondria produce energy, adenosine triphosphate (ATP), which is used as a major source of chemical energy. Mitochondria are also involved in pyruvate and fatty acid oxidation, nitrogen metabolism, and oxidative phosphorylation (Scarpulla, 2008). A unique feature of mitochondria is that they contain their own genetic system, mitochondrial DNA (mtDNA), which encodes a number of gene products that contribute to the structure and function of the organelle (Scarpulla, 2008). Although the mitochondrion has its own genetic information, the limited capacity of mtDNA requires nuclear genes to provide major respiratory subunits and to support mtDNA transcription and replication (Gopalakrishnan & Scarpulla, 1995). These nuclear genes encode most of the structural and catalytic components involved in the electron transport chain and oxidative phosphorylation system, as well as genes that control mitochondrial transcription, translation and DNA replication (Scarpulla, 2008). NRF1 is one of several key TFs that regulate transcription of nuclear genes encoding mitochondrial products. Other key TFs in nuclear control of mitochondrial biogenesis include GABP (GA binding protein transcription factor, alpha subunit, also known as NRF2), ESRRA (estrogen-related receptor alpha, also known as ERRα), PPARs (peroxisome proliferator-activated receptors), CREB1 (cAMP responsive element binding protein 1, also known as CREB), Sp1 and YY1 (Yin Yang 1, Scarpulla, 2008). As noted above, NRF1 was first identified and cloned in studies on transcriptional regulation of cytochrome *c* and cytochrome oxidase subunit genes (Evans & Scarpulla, 1989; Gopalakrishnan & Scarpulla, 1994; Virbasius *et al*., 1993). Bioinformatics analysis and chromatin immunoprecipitation (ChIP) using cancer cell lines showed that NRF1 binds to the promoters of many nuclear genes that are

involved in mitochondrial biogenesis and regulates transcription of these genes, including *TFAM* (mitochondrial transcription factor A), and others (Cam *et al*., 2004; Scarpulla, 2006).

Mitochondrial function is hypothesized to be associated with exercise and aging, NRF1 and its coactivators (e.g., PPARGC1A, also known as PGC-1α) have also been studied in these important physical conditions (Scarpulla, 2008). Under the normal situation in some cell types, NRF1 function involves co-activation by PPARGC1A to recruit RNA polymerase to promoters of nuclear target genes. It has been shown that exercise induces an increase in PPARGC1A and NRF1 in skeletal muscle as well as mitochondrial biogenesis (Murakami *et al*., 1998; Baar *et al*., 2002). In aging, reports have shown PPARGC1A and NRF1 expression levels are different comparing young versus aged subjects (Lezza *et al*., 2001, Viña *et al*., 2009), suggesting changes in mitochondrial biogenesis, but the answers are not yet definitive and the mechanism underneath is still obscure.

Studies of the function of NRF1, its coactivators, and other interacting TFs and regulators will help us better understand the mechanism of mitochondrial biogenesis, oxidative stress and aging, as well as how physical exercise can improve the activity of mitochondria to prevent its impairment by aging.

1.1.3 NRF1 in neuronal development

Although most of the early studies on NRF1 focused on mitochondrial functions, it is becoming increasingly clear that NRF1 not only regulates nuclear genes in mitochondrial biogenesis but also regulates numerous genes in neuronal and other systems.

For example, NRF1 was found to up-regulate transcription of the human *IAP* (Intergrin-associated protein or CD47) gene, involved with processes of memory formation in the rat hippocampus and regulation of dendritic outgrowth and synaptic transmission in developing

cortical neurons through the activation of MAPK (Huang, *et al*., 1998; Chang & Huang, 2004; Numakawa, *et al*., 2004). Subsequently, NRF1 was found to increase neurite outgrowth in part through the regulation of *IAP* (Chang, *et al*., 2005). NRF1 is also important for expression of *FMR1*, the fragile X mental retardation syndrome gene (Smith *et al*., 2004). More recent studies showed that NRF1 regulates gene expression of Synapsin I, a phosphoprotein that plays a role in regulation of axonogenesis and synaptogenesis (Wang *et al*., 2009), as well as several NMDA receptor (NR) genes which play important roles in neuronal plasticity and synaptic transmission associated neuronal functions (Dhar & Wong-Riley, 2009). The same group also investigated NRF1 regulation of cytochrome *c* oxidase (*COX*) subunits (Dhar *et al*., 2008; Dhar *et al*., 2009) to perhaps provide a link between energy generation and synaptic transmission. Additionally, studies of the NRF1 ortholog in zebrafish showed that loss of function led to a specific loss of photoreceptor neurons and their precursors (Becker *et al*., 1998) and expression of EWG, the *Drosophila* ortholog, is neuronal specific (Haussmann *et al*., 2008).

The above experimental data implies that NRF1 may play additional roles in neuronal systems and could be involved in neural-related disease pathways. My major focus has been to identify which systems are regulated by NRF1 and the mechanisms of action by NRF1.

1.1.4 Genome-wide studies of NRF1 binding sites

Transcription is initiated or repressed by the assembly of TFs at their binding sites in *cis*-acting DNA elements, usually in the promoter region of a gene, with the TSS defining where the gene begins to be transcribed into RNA (Wasserman & Sandelin, 2004). Besides promoters, other *cis*-acting DNA elements include enhancers (activation elements), silencers (repression elements) and insulator (boundary elements, Villard, 2004; Riethoven, 2010). These elements can be hundreds

or thousands of base pairs from 5' or 3' of the TSS (Villard, 2004; Riethoven, 2010). According to their function, TFs can most simply be divided into two groups. One group is the basal TFs that are necessary for transcription to occur, e.g., RNA polymerase II, and the other group includes the gene specific TFs which promote (as an activator) or suppress (as a repressor) basal transcription (Villard, 2004). Studying the exact mechanisms of how a TF regulates gene expression will be helpful in understanding gene functions and gene regulatory networks (GRNs), and how these operate in normal development as well as the roles in disease.

One of the approaches to understand the functions of a TF is to identify the target genes and regulated gene networks. In finding such a GRN, one can use the data to analyze TF function in certain diseases and identify new genes and mechanisms in disease regulation. In order to search for candidate TF binding sites in the genome, a suitable first step in the era of genome sequences is a computational approach for specific binding site sequences to identify candidate genes. Next, several genome-wide experimental methods have been established to prove functional TF binding sites: 1) microarray analysis, e.g., Chromatin immunoprecipitation on chip (ChIP-chip); 2) gene silencing by RNA interference (siRNA) or TF overexpression (a complementary approach to siRNA); and 3) next generation sequencing, e.g., ChIP-seq and RNA-seq (Goutsias & Lee, 2007; MacQuarrie *et al*., 2011; Pareek *et al*., 2011). With these approaches, a list of target genes will be identified and can be focused on for more in depth functional studies. To date, many experimental methods, including *in vitro* assays such as electrophoretic mobility shift assay (EMSA) (Fried & Crothers, 1981), reporter assays in cell culture, and *in vivo* methods such as ChIP (Kuo & Allis, 1999) have been applied to elucidate the interaction of a TF and DNA sequences. Nevertheless, these studies can only be performed when the location of regulatory element is known or predicted. In large and complex genomes, discovery of *bona fide* TF binding sites are challenging due to

factors such as 1) poor definition of gene promoters; 2) the distance between TF binding sites (e.g., in enhancers) and promoters can be 1 kb or more and up to a megabase; 3) the degeneracy in TF recognition motifs around a consensus nucleotide sequence; and 4) the variety of mechanisms, such as spacing and arrangement of TF binding sites, that are involved in transcriptional regulation (Hallikas *et al*., 2006; Villard, 2004).

Recent approaches have used computational bioinformatics to develop algorithms capable of searching genome sequences for potential TSS and TF binding sites. Typically, several classes of computational programs have been applied to the search for potential TF binding motifs in a genome and each has strengths and weaknesses (Wasserman & Sandelin, 2004). With the availability of representative mammalian genome sequences, comparative genomics provides a powerful tool to systematically discover functional elements in a genome (Wasserman *et al*., 2000; Dubchak *et al*., 2000; Xie *et al*., 2005). In particular, the NRF1 motif has been found to be the most highly conserved putative TF binding motif in the human genome (Xie *et al.*, 2005). The basis for this is likely to have several explanations including 1) NRF1 has a longer recognition motif than most TFs; 2) the NRF1 binding site allows less degeneracy than most TF binding sites which can be highly degenerate; and 3) NRF1 gene regulation, once established by evolutionary processes, may be associated with stronger selective pressure and/or reduced compensatory changes in TF regulation than for other TF-target gene interactions. The first two of these explanations are involved while the third is a hypothesis that remains to be tested. Experimentally, these considerations provide a strong rationale for inclusion of bioinformatics studies to identify and understand NRF1 functions in GRNs.

To date, genome-wide studies in human by ChIP-chip found NRF1 binding sites in 5.3% of ~13,000 gene promoters (Cam *et al*., 2004). Nevertheless, this study used a karyotypically highly

rearranged cancer cell line, and thus many epigenetically regulated or tissue-specific target genes would not have been identified (including virtually all the target genes identified in this thesis). A strictly bioinformatics study examining the distribution of 8-mer nucleotide sequences (FitzGerald *et al.*, 2004) found that a sequence (CGCVTGCG) similar to NRF1 binding sites was found in 6% of the 13,010 promoters analyzed, although the result may be slightly overestimated since the core of the consensus NRF1 binding site is longer (GCGCANGCGC). Indeed, a GCGCatGCGC motif was found in 5.3% of 10,577 promoters (Xi *et al*., 2007). Using comparative analysis of the human, mouse, rat and dog genome sequences to identify conserved regulatory motifs in promoters of ~ 17,700 genes, Xie *et al.* (2005) found 6.1% had conserved NRF1 binding sites. These studies provide a preliminary estimate of the frequency of NRF1 target genes in the human genome, although as no experimental confirmation was done it is unknown what percentage of identified sites are functionally relevant. Furthermore, none of these studies examined non-promoter sequences, so the potential contribution of NRF1 to long-range gene regulation is completely unknown.

1.2 PRADER-WILLI SYNDROME

1.2.1 Clinical phenotype and diagnosis

Prader-Willi Syndrome (PWS) is a complex congenital disorder that involves genetic abnormalities in chromosome 15q11-q13, with an incidence of ~1:15,000-1:25:000 births (Holm *et al*., 1993; Buiting, 2010; Wattendorf & Muenke, 2005). It is the most common form of obesity caused by a genetic syndrome (Wattendorf & Muenke, 2005). The clinical phenotype of PWS includes severe hypotonia, respiratory distress, feeding difficulties and failure-to-thrive in early infancy, followed by hyperphagia and development of obesity in later infancy or early childhood (Holm *et al*., 1993; Nicholls & Knepper, 2001; Gunay-Aygun *et al*., 2001). Patients are delayed in early motor and language development, as well as having some degree of cognitive impairment and hypogonadism (Holm *et al*., 1993; Gunay-Aygun *et al*., 2001). Other common features include short stature, characteristic facial features, strabismus and scoliosis (Holm *et al*., 1993; Wattendorf & Muenke, 2005). Patients also have common behavioral abnormalities, e.g., temper tantrums, stubbornness, manipulative behavior and obsessive-compulsive characteristics (Holm *et al*., 1993; Nicholls & Knepper, 2001). A prevalence rate of 25% for non-insulin-dependent diabetes mellitus was found in PWS adults due to obesity (Butler *et al*., 2002).

1.2.2 Genetics and epigenetics

At the molecular level, genes within the 2 Mb PWS region are regulated by the process of genomic imprinting. Genomic imprinting is an epigenetic mechanism leading to uniparental expression of certain genes in mammals and is established by a genetic element called the imprinting center (IC) during gametic and early preimplantation development (Nicholls & Knepper, 2001; Morison *et al*., 2005). The most common cause of PWS is a 5-6 Mb *de novo* deletion exclusively on the paternally-inherited allele. There are two classes of deletions in PWS patients, with two proximal breakpoints 1 and 2 (BP1, BP2) in 15q11 and a single distal BP3 in 15q13. PWS patients can also have maternal uniparental disomy (UPD) of chromosome 15, or in rare cases imprinting defects generating a maternal imprint on the paternally-derived chromosome (Nicholls & Knepper, 2001; Horsthemke & Buiting, 2006). In each circumstance, PWS arises from the functional loss of a set of paternally-expressed genes.

To date, at least 13 paternally-expressed genetic loci have been identified in the PWS-region in human, six of which encode polypeptides, including *MKRN3* (Jong *et al*., 1999a; 1999b), *MAGEL2* (Boccaccio *et al*., 1999; Lee *et al*., 2000), *NDN* (Jay *et al*., 1997; MacDonald & Wevrick 1997), *C15orf2* (Buiting *et al*., 2007), and the bicistronic *SNURF-SNRPN* [Gray *et al*., 1999a, **Figure 1.2(a)**]. A second imprinted disorder, Angelman syndrome (AS), is a neurological disease associated with genetic mechanisms that lead to a loss of function of a maternally-expressed imprinted gene, *UBE3A*, located telomeric to the PWS domain [**Figure 1.2(a)**]. Most PWS genes have an exclusive or predominant expression pattern in neurons (Cavaillé *et al*., 2000; Nicholls & Knepper, 2001; Runte *et al*., 2001), and their functions are only partially understood. The bicistronic locus *SNURF-SNRPN* encodes two polypeptides. *SNRPN* encodes a brain-specific

spliceosome protein, SmN, which replaces the otherwise constitutive SmB'/B protein during late fetal and postnatal brain development (Sharpe *et al*., 1990, McAllister *et al*., 1989), while the upstream *SNURF* encodes an independent nuclear protein of unknown function (Gray *et al*., 1999a). *SNURF-SNRPN* is also the host gene of five classes of box C/D small nucleolar RNAs (snoRNAs). The snoRNAs include *SNORD107* (previously *HBII-436*, 1 copy), *SNORD64* (previously *HBII-13*, 1 copy), *SNORD109A* (previously *HBII-438A*, 1 copy), *SNORD116* (previously *HBII-85*, 29 copies), *SNORD115* (previously *HBII-52*, 47 copies) and *SNORD109B* (previously *HBI-438B*, 1 copy, Cavaillé *et al*., 2000; Runte *et al*., 2001; Buiting, 2010; Zhu, *et al*., in preparation), most of which have unclear function and targets, except that *SNORD115* has been identified to regulate serotonin receptor 2C mRNA splicing (Kishore & Stamn, 2006).

The complex genetic locus encoding the polycistronic *SNURF-SNRPN*-snoRNA loci appears to play a major role in PWS, based on the location of the imprinting regulatory region at its 5' end [**Figure 1.2(a)**; Nicholls & Knepper 2001; Horsthemke & Buiting 2006], its disruption by two classes of balanced translocations (Wirth *et al*., 2001), and the finding that mice with a deletion of *Snurf-Snrpn* and the snoRNA clusters have a "partial mouse PWS" phenotype (Tsai *et al*. 1999). More recent studies showed that NECDIN (a MAGE-family protein, encoded by *NDN*) and MAGEL2 bind to several proteins in the region of the centrosome, and that axonal outgrowth is abnormal in *Ndn*-deficient embryos (Lee *et al*., 2005), while other studies suggest that in preadipocytes NECDIN inhibits a TF, E2F4, to prevent adipogenesis which is consistent with its long-postulated role in post-mitotic cells (Tseng *et al*., 2005). NECDIN also appears to play a role in the neuronal function of respiration (Ren *et al*., 2003; Zanella *et al*., 2008). *Magel2* is a clock-controlled circadian output gene (Kozlov *et al*., 2007). All PWS genes have similar expression patterns with highest or exclusive expression in neurons, especially during development (Cavaillé *et*

al. 2000; Runte et al., 2001; Lee *et al*., 2003; Watrin *et al*., 2005), and a common regulatory element specific for neurons has been proposed (Watrin *et al*., 2005).

The PWS region has an intriguing evolutionary history in mammals and shows clear evidence of relatively rapid and recent evolutionary change. *SNRPN* and the PWS region snoRNAs derive from an ancestral duplication of a *SNRPB'* gene in a therian ancestor (Gray *et al*., 1999; Rapkins *et al*., 2006; unpublished data). *Mrkn3*, *Magel2*, *Ndn*, and *Frat3* in rodents are all intronless genes and acquired by retrotranposition events in a eutherian or rodent ancestor, respectively (Jong *et al*., 1999a; 1999b; Gray *et al*., 2000, 2001; Chai *et al*., 2001). *C15orf2* shows a predominantly testis-specific expression pattern (Färber *et al*., 2000), with low expression in fetal brain and originates from a retrovirus in an ancestral primate (Buiting *et al*., 2007). In mouse, the homologous PWS region is in chromosome 7C [**Figure 1.2(b)**], except that mouse doesn't have *C15orf2* and *PWRN1* (Buiting *et al*., 2007), but gained *Frat3* by retrotransposition (Chai *et al*., 2001). The rodent locus also does not have *Snord109* copies and has a rodent specific microRNA (miRNA), miR-344 [**Figure 1.2(b)**]. Loss of imprinted gene expression in mouse 7C leads to a similar but more severe neonatal phenotype as human (Nicholls & Knepper, 2001; Stefan *et al*., 2005; Stefan *et al*., in preparation).

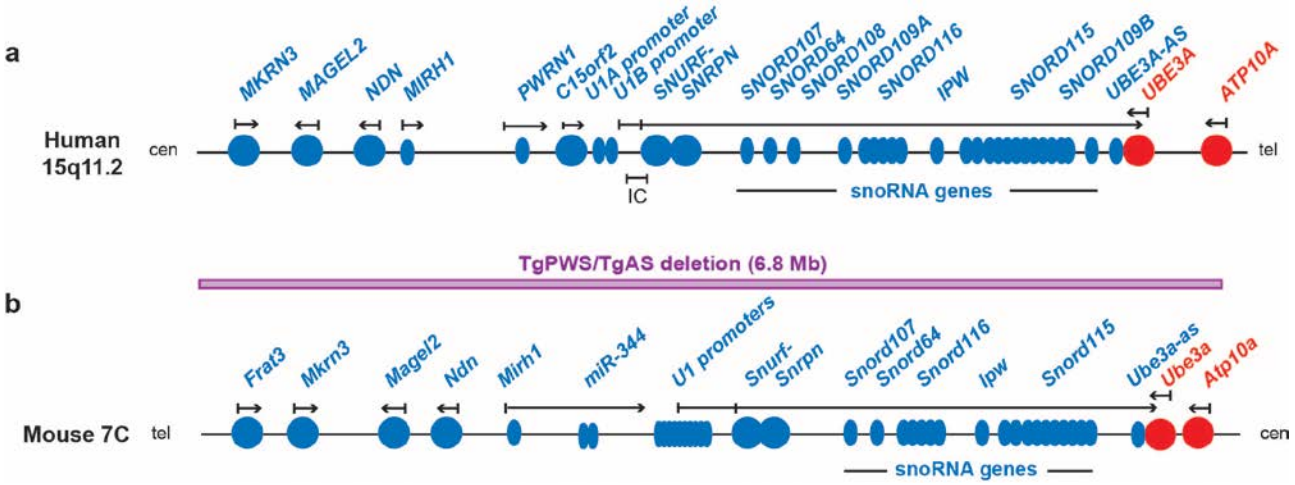


Figure 1.2. Prader-Willi syndrome (PWS) imprinted domain in (a) human chromosome 15q11.2 and (b) mouse chromosome 7C. Genes in blue are paternally expressed, while genes in red are maternally expressed. Circles indicate genes encoding proteins, ovals indicate RNA genes, and arrows indicate the direction of gene transcription. cen, centromere; tel, telomere; IC, imprinting center. Modified from Stefan, *et al*., in preparation.

1.3 CIRCADIAN RHYTHMS

As the word “circadian” was first derived from the latin word “circa diem”, it means “about a day” (Roenneberg & Merrow, 2005; Toh, 2008). Circadian rhythms exist in virtually all living organisms and function as an internal body clock that regulates the metabolic, physical and behavioral paces of the body (Panda *et al*., 2002; Roenneberg & Merrow, 2005; Toh, 2008). The fundamental trigger of circadian rhythms is environmental cues, such as light and temperature (Roenneberg & Merrow, 2005). It is hypothesized that circadian oscillators are an advantage for living organisms to adjust internal physiology to adapt to environmental changes (Panda *et al*., 2002).

1.3.1 Physiology and behavior

The daily sunrise and sunset are considered to be the ultimate force for circadian rhythms (Roenneberg & Merrow, 2005). The central mammalian circadian pacemaker resides in the suprachiasmatic nucleus (SCN) of the hypothalamus (Maywood *et al*., 2007; Takahashi *et al*., 2008). The SCN is composed of bilateral nuclei, each comprising approximately 10,000 neurons which are autonomous cells (Takahashi *et al*., 2008).When these neurons in the SCN are entrained by projection from retinal neurons, they use neuronal and hormonal mechanisms to coordinate rhythmic physiology and behavior. Studies have shown that damage to the SCN eliminates endogenous circadian rhythms, which supports its role as circadian pacemaker (Inouye & Kawamura, 1979). In addition to the SCN, peripheral organs, such as liver and heart, have circadian

rhythms that receive neural and hormonal output signals from the SCN (Rutter *et al*., 2002; Hastings *et al*., 2007) and additionally can also be entrained by food or energy balance (Mendoza, 2007).

In mammals, circadian rhythms influence many aspects of physiology and behavior, including sleep-wake cycles, energy metabolism, hormonal regulation, heart rate, blood pressure and body temperature (Triqueneaux *et al*., 2004; Gallego & Virshup, 2007). Therefore, it is reasonable to consider the goal of circadian rhythms is to optimize metabolism and energy balance for sustaining life of a living organism (Levi & Schibler, 2007).

1.3.2 Molecular clocks in human

In animals, the cellular circadian cycle is determined molecularly by oscillations of positive and negative transcriptional and post-transcriptional feedback regulation. In the positive feedback loop, CLOCK and BMAL1, two basic helix-loop-helix (bHLH)-PAS TFs, form a heterodimeric complex and function as a transcriptional activator in the circadian pacemaker to activate the transcription of circadian core genes, period (PER) and cryptochrome (CRY), as well as other circadian output genes, through E-box DNA binding elements (5'- CACGTG-3', Steeves *et al*., 1999; Gekakis *et al*., 1998; Kondratov *et al*., 2003). On the other hand, PER and CRY also heterodimerize and mediate the negative limb of a transcriptional feedback loop in the circadian system (Sun *et al*., 1997; Darlington *et al*., 1998; Hastings *et al*., 2007; Brown *et al*., 2005). In this feedback system, the CLOCK/BMAL1 complex binds to E-boxes in the promoters of the *PER* and *CRY* genes and stimulates transcription, but as the concentration of PER/CRY increases, they inhibit CLOCK/BMAL1 function. Subsequently, PER and CRY are degraded and cycling of this system generates the circadian oscillations (Reppert & Weaver, 2002). In addition, *NR1D1* (also

known as *REV-ERBα*) forms an ancillary feedback loop by being a repressor of *BMAL1* transcription (Takahashi *et al*., 2008). Another circadian regulator is *DBP*, which encodes a TF that regulates many clock output genes as well as *PER1* (Yamaguchi *et al*., 2000).

Post-translational modifications such as phosphorylation, ubiquitination, sumoylation, and acetylation have been shown to play important roles in the regulation of circadian expression (Cardone, *et al*., 2005; Takahashi *et al*., 2008). Circadian timing of PERs is controlled through progressive phosphorylation by kinases, CSNK1A1 (also known as CKIα, Hirota *et al*., 2010), CSNK1D (also known as CKIδ, Lee *et al*., 2009) and CSNK1E (also known as CKIε, Akashi *et al*., 2002; Takahashi *et al*., 2008; Lee *et al*., 2009). The stability and cellular localization of PER is also regulated by phosphorylation (Akashi *et al*., 2002; Blau, 2008). For CRY1, phosphorylation and protein stability are rhythmically regulated by AMPK, an adenosine monophosphate (AMP)-activated protein kinase (Lamia *et al*, 2009). BTRC (also known as β–TrCP1) and FBXL3, both F-box proteins and subunits of ubiquitin protein ligase complexs, can target phosphorylated PER and CRY proteins, respectively, and promote polyubiquitylation for the 26S proteosomal degradation pathway (Reischl *et al*., 2007; Busino *et al*., 2007). In the circadian system, sumoylation was first found in a lysine residue (Lys^{259}) of BMAL1 by SUMO1 to enhance the circadian expression of BMAL1 (Cardone *et al*., 2005). Such a process is induced by CLOCK (Cardone *et al*., 2005). In addition, CLOCK protein can rhythmically acetylate BMAL1; subsequently, acetylation of BMAL1 facilitates recruitment of CRY1 to CLOCK/BMAL1 for transcription repression (Hirayama *et al*., 2007). SIRT1 (histone deacetylase sirtuin 1), on the other hand, can interact with CLOCK to deacetylate BMAL1 and PER2 (Nakahata *et al*., 2008; Asher *et al*., 2008).

Additionally, histone acetylation and methylation are also crucial in circadian regulation (Takahashi *et al*., 2008; Katada & Sassone-Corsi, 2010). Rhythmic expression of mouse *Per1* and

Per2 are regulated by histone H3 acetylation and deacetylation (Etchegaray *et al*., 2003; Curtis *et al*., 2004; Naruse *et al*., 2004). Similarly, rhythmic histone acetylation was reported in circadian transcription of *Dbp* by the binding of CLOCK/BMAL1 to an intronic *Dbp* enhancer (Ripperger & Schibler, 2006). Later studies showed that CLOCK protein has histone acetyltransferase (HAT) activity (Doi *et al*., 2006; Hirayama *et al*., 2007), while SIRT1 is a NAD+-dependendent deacetylase to histone H3 (Nakahata *et al*., 2008; Asher *et al*., 2008). Circadian expression of SIRT1 was found in mouse liver and fibroblast cells with a constant level of *Sirt1* mRNA, indicating posttranscriptional regulation involved in circadian regulation of SIRT1 (Asher *et al*., 2008). Since the function of SIRT1 depends on NAD+, it provides a link between cellular metabolism and circadian rhythms. Finally, recent studies indicated that a chromatin activator MLL1 is part of the CLOCK/BMAL1 complex and plays a crucial role in circadian transcription and H3K4 trimethylation (Katada & Sassone-Corsi, 2010). Overall, these findings in post-translational regulation and histone modification for chromatin remodeling change the traditional view of the control of molecular clock purely by a transcriptional feedback loop. The network of transcriptional and post-transcriptional regulation shows how complex and hierarchical the molecular control of circadian regulation is. Together, these processes create the dynamic changes of chromatin structure in circadian expression and thus reflected by rhythmic changes on mRNA and protein levels of circadian input and output genes.

During the past decade, studies in circadian rhythms have made remarkable progress. One piece of evidence is that many new genes have been studied and found to play roles in circadian regulation (Asher *et al*., 2008; Lee *et al*., 2009; Katada & Sassone-Corsi, 2010, Hirota *et al*., 2010). However, more evidence indicated that new regulatory genes may involve different levels of regulation to fine-tune circadian regulation (Bozek *et al*., 2010).

1.3.3 Circadian rhythms and disease

Circadian rhythms direct neural and hormonal control of daily activity, feeding, and sleep cycles. They are disrupted by travel, shift-work, and in disease including sleep, eating and psychiatric disorders, obesity, diabetes, cardiovascular disease, infection, and cancer (Kondratov & Antoch, 2007; Weldemichael & Grossberg, 2010; Wulff *et al*., 2010; Gravotta *et al*., 2011).

Since circadian rhythms are highly correlated to body metabolism, many studies showed association between circadian clock and metabolic abnormalities (Bass & Takahashi, 2010; Huang *et al*., 2011). Glucose concentrations in the blood are highly rhythmic (Bass & Takahashi, 2010). Patients with type II diabetes have abnormal circadian insulin secretion (Boden *et al*., 1999). *Clock* and *Bmal1* mutant mice have impaired glucose tolerance, reduced insulin secretion and other phenotypes that lead to the development of diabetes mellitus and obesity (Turek *et al*., 2005; Marcheva *et al*., 2010).

Another common feature for disease associated with circadian rhythms is an abnormality in sleep. The timing of sleep is controlled by the circadian clock with the release of melatonin and a reduction of body temperature. It has been reported that more than 80% of patients with depression or schizophrenia have sleep complaints (Wulff *et al*., 2010). Decrease of melatonin secretion is often found in patients with Alzheimer's disease and phenotypic features of day time agitation, night-time insomnia and restlessness (Weldemichael & Grossberg, 2010; Wulff *et al*., 2010). Similarly, for patients with Parkinson's disease, it is common to have night-time sleep disturbances and daytime sleepiness (Wulff *et al*., 2010).

At the molecular level, the first disorder that linked directly to core circadian genes is familial advanced sleep phase syndrome (FASPS), which has missense mutation in *PER2*

(Takahashi *et al*., 2008). Insomnia is one of the most common sleep disorders and affects approximately 10-15% of the adult population (Ban *et al*., 2011). A cohort study has linked insomnia to SNPs in *ROR1* and *PLCB1*, genes that have previously been associated with bipolar disease and schizophrenia, respectively (Ban *et al*., 2011). A mutation has been found in *DEC2*, encoding a repressor to the CLOCK/BMAL1 complex, in a family-based study of lifelong shorter daily sleep time (He *et al*., 2009). Subjects with the *DEC2* mutation have a short sleep phenotype (He *et al*., 2009).

Interestingly, clinical studies show that disease treatment is often sensitive to the time of administration (Levi & Schibler, 2007). Consequently, rhythmic processes in the body should be considered in personalized pharmacotherapy for more effective treatment (Ohdo, 2010), including in cancer treatment (Kondratov & Antoch, 2007). Therefore, a better understanding of how body clocks work will not only provide insights into the pathogenesis of these and other common disorders, but also improve treatments for disease.

2.0 TRANSCRIPTIONAL AND EVOLUTIONARY MECHANISMS IN PRADER-WILLI SYNDROME

2.1 INTRODUCTION

2.1.1 Regulation of the PWS imprinted domain in somatic cells

As described in **Chapter 1.2.**, PWS is an inborn disease that involves chromosomal abnormalities of human chromosome 15q11.2. In the 2 Mb human PWS domain (**Figure 2.1)**, our laboratory has identified 14 NRF1 binding sites in the *SNURF-SNRPN* promoter (2 sites), *SNURF-SNRPN* intron 1 enhancer (2 sites), a CpG-islet site 3.7 kb 5' of *MKRN3* (intergenic in mouse), the *NDN* promoter, the *U1A* and *U1B* promoters, putative enhancer elements upstream of *U1A* and *U1B*, and in an element termed the "NRF1 cluster". The "NRF1 cluster" element is located between a cluster of intronless genes in proximal 15q11.2 and the *U1-SNURF-SNRPN-snoRNA* region and consists of four conserved NRF1 binding motifs (**Figure 2.2**, Stefan *et al*., in preparation). The stereospecific arrangement of 4 clustered NRF1 motifs suggests binding to the same face of the α-helix and potential for cooperative interactions similar to known enhancer elements (Stefan *et al*., in preparation). Studies by our laboratory also found that the "NRF1 cluster" element shows brain-specific DNA methylation, and has strong enhancer function in reporter assays (Stefan *et al*., in preparation). The "NRF1 cluster" element also shows NRF1 binding and an enhancer-pattern

(Heintzman *et al*., 2007; Schübeler, 2007) of histone H3K4 methylation based on ChIP studies (Stefan *et al*., in preparation). Similarly, in the mouse orthologous region, chromosome 7C, we found 18 NRF1 *cis*-binding sites in 11 gene promoters, two enhancers, and the "NRF1 cluster", all differentially methylated and bound *in vivo* by NRF1 within active chromatin [acetylated histone H4 and dimethylated histone H3 (Lys4)] on the paternal allele only (Stefan *et al*., in preparation). Combined, the human and mouse studies indicated that NRF1 is a "master" regulator of the PWS region (Stefan *et al*., in preparation). However, the previous human studies used SK-N-SH cells which do not express *MAGEL2*, *PWRN1*, or *C15orf2*, and only express a low level of *MIRH1* (Stefan *et al*., in preparation).

In order to confirm and extend analysis of the role of NRF1 in regulating PWS genes, my study has used ChIP assays to examine NRF1 binding in the GM11715 mouse/human somatic hybrid cell line. The GM11715 cell line was derived by microcell fusion of human diploid fetal lung fibroblasts with mouse A9 cells (Ning *et al*., 1992) and is *Hprt*- (Cos *et al.,* 1974), which can be used in drug selection by HAT for gene targeting (Capecchi, 2005). Similarly, A15 (and its derivative, A15-1, which has a truncated version of chromosome 15) is also a somatic cell hybrid derived from mouse A9 cells (McDaniel & Schultz, 1992) but has a more active, hypomethylated epigenetic state in 15q11.2 than does GM11715 (Gabriel *et al*., 1998), which provides alternative cell line(s) for such studies. Expression studies have found that all of the human PWS genes are expressed robustly in the GM11715 cell line with the exception of *MKRN3*; in contrast, the homologous mouse genes are not expressed (**Figure 2.3**). Because the three somatic cell hybrids all express apparently robust levels of PWS genes and contain only one paternally-derived human chromosome 15 (**Figure 2.3**, Gabriel *et al*., 1998), these provide an apparently ideal system for the

study of NRF1 regulation in the PWS domain by gene targeting of a single copy of the "NRF1 cluster".

Gene targeting is based on the principle of homologous recombination, and has been applied to many studies using mouse embryonic stem (ES) cell studies to examine gene function and phenotype, and a few studies in gene regulation (Sedivy & Dutriaux, 1999), as well as studies in somatic cells (Porter & Itzhaki, 1993). If the disruption of the "NRF1 cluster" by gene-targeting was successful, it would allow testing the hypothesis that this element has long-range *cis*-regulatory function in gene expression across the PWS domain. Importantly, this approach will allow analysis of gene expression and changes in chromosomal structure of the entire PWS domain in a chromosome setting.

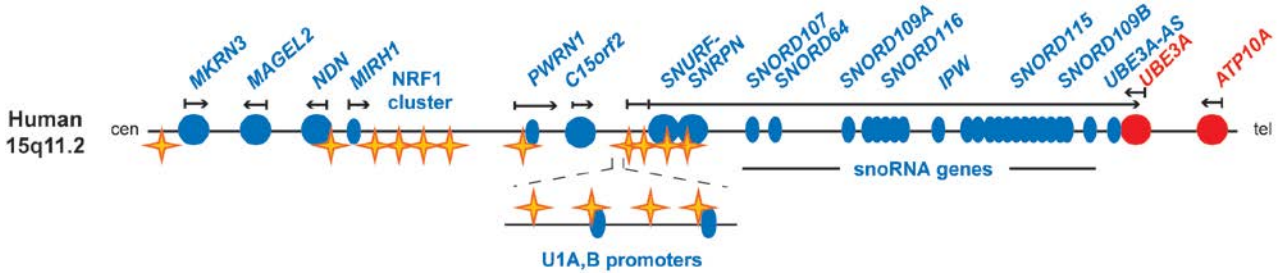


Figure 2.1. NRF1 is a master regulator of the human Prader-Willi syndrome (PWS) imprinted domain in chromosome 15q11.2. Genes in blue are paternally expressed, while genes in red are maternally expressed. Circles indicate genes encoding proteins, ovals indicate genes encoding small RNAs, and arrows indicate the direction of gene transcription. Gold stars indicate verified NRF1 binding sites. cen, centromere; tel, telomere; Modified from Stefan *et al*., in preparation.

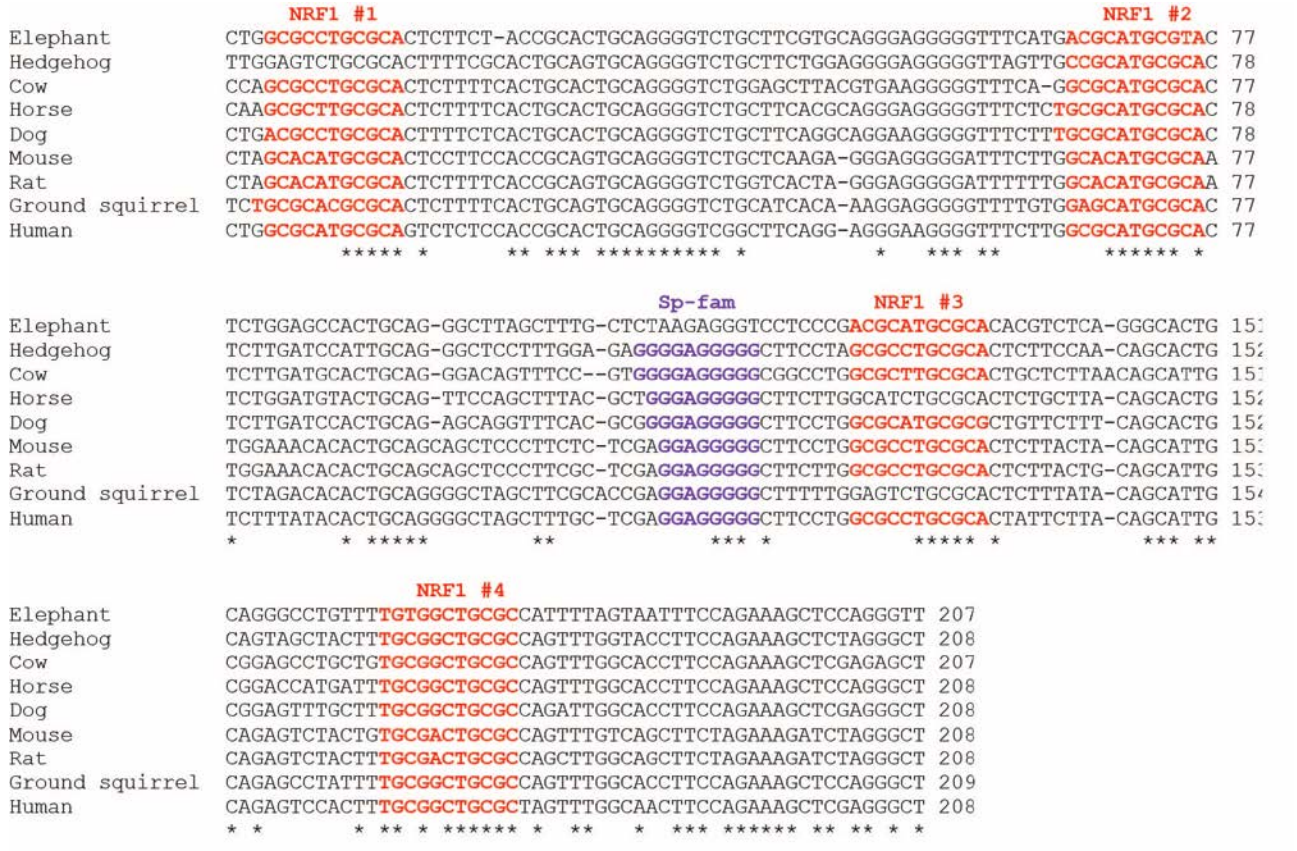


Figure 2.2. Multi-sequence alignment of the "NRF1 cluster" in mammals. NRF1 sites are highlighted in bold red. Sp-family site is highlighted in bold purple.

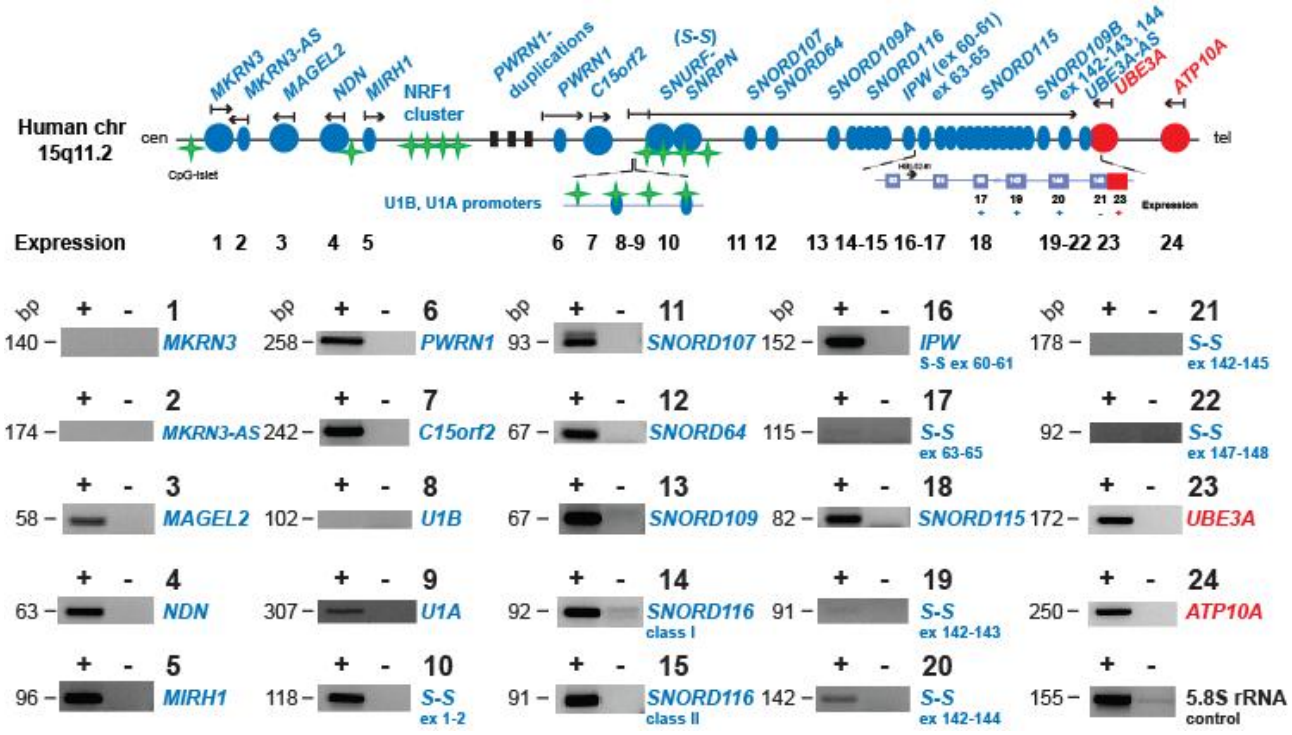


Figure 2.3. Expression studies on GM11715 cells. Abbreviations: "+": RT positive; "-": RT negative; 5.8S rRNA as a positive control (Li, S & R. D. Nicholls).

2.1.2 Genomic imprinting and disease

Genomic imprinting is a type of epigenetic modification that has roles in mammalian development, especially placental and fetal development, childhood and adult behavior (Hall, 1990; Li & Sasaki, 2011). It results in the differential expression of genetic material from the maternally- and paternally-derived alleles. Consequently, genomic imprinting does not show Mendelian inheritance and provides an explanation for some of the genetic events that do not follow single-gene inheritance (Hall, 1990).

Imprinting was first identified during nuclear transfer experiments in mice in the 1980's. By transplantation of either the haploid androgenetic or the gynogenetic nucleus to haploid recipient eggs, only the ones with each copy of paternal and maternal pronuclei resulted in development to term, suggesting the existence of imprinting mechanisms (Surani *et al*., 1986; Ciccone & Chen, 2009). The establishment and removal of genomic imprinting occurs primarily during gametogenesis (but is not completed until early preimplantation development) by epigenetic modification, including DNA methylation by DNA methyltranferases (DNMTs) and DNA demethylases (DMEs), and histone modification by lysine methyltransferases (KMTs) and lysine demethylases (KDMs) (Kota & Feil, 2010; Law & Jacobsen, 2010). Most imprinted genes have regulatory sequences that are differentially methylated regions (DMRs, Uribe-Lewis *et al*., 2011). Once genomic imprinting is established, it is then passed to the zygote through fertilization and maintained throughout somatic development and adult life (Wood & Oakey, 2006; Li & Sasaki, 2011). Only in the formation of primordial germ cells are such modifications erased and new imprints set for the next gametogenesis cycle (Li & Sasaki, 2011). As a result, imprinting maintains permanent genetic features within a generation, but may alter genetic features between generations. The first identified imprinted genes were the paternally expressed insulin-like growth factor 2 (*Igf2*, DeChiara *et al*., 1991) and the maternally expressed insulin-like growth factor type-2 receptor (*Igf2r,* Barlow *et al*., 1991) in mice. To date, 41 transcriptional units (TUs) in human and 71 TUs in mouse have been reported to be imprinted, in which 29 TUs are shared by human and mouse (Morison *et al*., 2001, http://igc.otago.ac.nz/home.html). Among these TU, 30% (25/83) of them encode noncoding RNAs.

Defects in genomic imprinting are associated with different types of disease. Uniparental disomy (UPD) is a condition in which an individual inherited both copies of a chromosome from

one parent only (Yamazawa *et al*., 2010). Usually UPD is not associated with a phenotype unless imprinted genes are present on the chromosome with UPD (or, if UPD results in homozygosity for a recessive mutation). Lack of either paternal or maternal imprinted genes in UPD causes certain inborn diseases. So far, nine types of UPD (Morison *et al*., 2001; Hirasawa & Feil, 2010; Yamazawa *et al*., 2010) have been reported, among which four types are maternal UPD, including Russell-Silver syndrome (or Silver-Russell syndrome, chromosome 7p11-13 and 7q31-qter, Hannula *et al*., 2001; Eggermann, 2010), maternal UPD14 syndrome (MUPD14 syndrome, 14q, Kayashima *et al*, 2002; Ogata *et al*., 2008), PWS (15q11-13, Nicholls *et al*., 1989; Nicholls & Knepper, 2001) and maternal UPD20 (MUPD 20, only three cases, among which one is mosaic trisomy 20 (Chudoba *et al*., 1999; Eggermann *et al*., 2001; Salafsky *et al*., 2001). In addition, five types have paternal UPD, including transient neonatal diabetes mellitus (6q24, Shield, 2000; Aviram *et al*., 2008), Beckwith-Wiedemann syndrome (BWS, 11p15, Smith *et al*., 2007), paternal UPD14 syndrome (14q32, Ogata *et al*., 2008; Yamazawa *et al*., 2010), Angelman syndrome (AS, 15q11-13, Robinson *et al*., 2000; Nicholls & Knepper, 2001) and pseudohypoparathyroidism type 1b (20q13, Morison *et al*., 2001; Amor & Halliday, 2008; Fernández-Rebollo *et al*., 2010). Imprinting abnormalities have also been associated with cancers, which have characteristics of global hypomethylation and site-specific hypermethylation, as well as loss of imprinting in certain types of cancer (Feinberg, 2007). For examples, congenital loss of imprinting at the *IGF2-H19* locus occurs in Wilms tumor and BWS individuals have an increased risk for Wilms tumor and hepatoblastoma (Uribe-Lewis *et al*., 2011). Therefore, a better understanding of the mechanisms of genomic imprinting will not only improve the current functional studies of imprinted genes during gametogenesis and embryonic cell development (Li & Sasaki, 2011), but also improve understanding of mechanisms for some complex diseases.

2.1.3 Evolutionary insights and relationships of the PWS *SNRPN* imprinting locus and snoRNAs

Genomic imprinting is widespread in placental mammals (eutherian mammals) and also occurs in marsupials (metatherian mammals, Renfree *et al*., 2009). These two branches of mammals diverged from a common ancestor about 173-190 million years ago (mya, Kumar & Hedges, 1998; van Rheede *et al*., 2006). The first complete genome sequence of a metatherian, *Monodelphis domestica* (gray short-tailed opossum), revealed that non-protein-coding sequences between metatherians and eutherians were up to 80% conserved (Mikkelsen *et al*., 2007) indicating the high sequence conservation in the gene regulatory elements and their origins from a common ancestor. The study of the metatherian genome provides great insights in understanding the evolutionary history of human genes. As a relatively young genetic mechanism in evolutionary terms, imprinting in mammals has been found to be acquired by a combination of ancient gene functions present in non-mammalian vertebrates and evolutionary young gene loci without clear lower vertebrate ancestors. It has been shown that chicken doesn't have imprinted genes (O'Neill *et al*., 2000). Metatherian imprinted genes include *IGF2* (O'Neill *et al*., 2000; Lawton *et al*., 2008), *PEG1/MEST* (Suzuki *et al*., 2005), *IGF2R* (Killian *et al*., 2001) although it is hypomethylated compared to mouse *Igf2r* (Weidman *et al*., 2006), and *PEG10* (Suzuki *et al*., 2007). The unique and expanded stage of postnatal embryonic development in metatherians provides a good model to study the origin of imprinting from non-placental mammals to placental mammals, since the development of the placenta in eutherians has been suggested to relate to the origin of genomic imprinting (Suzuki *et al*., 2007).

PWS is a genomic imprinting disorder that occurs in approximately 1 per 15,000 newborns. The clinical features of PWS include neonatal failure-to-thrive with childhood-onset severe obesity and other endocrine and behavioral abnormalities. PWS arises from functional loss of multiple imprinted, paternally-expressed genes, which are located on human chromosome 15q11.2 (Nicholls & Knepper, 2001). So far, there are at least thirteen paternally expressed genes that have been identified, including five protein coding genes, *MKRN3*, *MAGEL2*, *NDN*, and a bicistronic *SNURF-SNRPN*. The bicistronic locus *SNURF-SNRPN* normally encodes two polypeptides: *SNRPN* encodes a brain-specific spliceosomal protein, SmN, which replaces SmB'/B protein during postnatal brain development (Sharpe *et al*., 1990; McAllister *et al*., 1989), while the upstream *SNURF* encodes an independent nuclear protein with an unclear function (Gray *et al*., 1999a). *MKRN3*, *MAGEL2* and *NDN* are all intronless genes (Chai *et al*., 2001) and their function remains largely unknown. In mouse, the homologous PWS region is on chromosome 7C, except that mouse doesn't have the testis-specific *C15orf2* (Färber *et al*., 2000), but gained *Frat3* by retrotransposition (Chai *et al*., 2001). The loss of imprinted gene expression on mouse 7C showed similar neonatal defects as in human (Stefan *et al*., 2005). In the PWS gene region, *SNURF-SNRPN* also serves as the host of five classes of C/D-box small nucleolar RNAs (snoRNAs), *SNORD107*, *SNORD64*, *SNORD109*, *SNORD116* and *SNORD115* (Cavaillé *et al*., 2000; Runte *et al*., 2001). SnoRNA is a group of RNA that guides chemical modification of ribosomal RNA (rRNA) or other RNAs. There are two classes of snoRNAs, C/D-box snoRNAs and H/ACA-box snoRNAs. Named by conserved sequence motifs box C/C' (5'-CUGA-3') and D/D' (5'-RUGAUGA-3'), the C/D-box snoRNAs contain one or two 10-21 nucleotide long antisense elements 5' of box C/C', which are complementary to the mature ribosomal RNA and direct site-specific 2'-*O*-methylation (Tyc & Steitz, 1989; Kiss-Laszlo *et al*., 1996; 1998). The C/D-box snoRNAs can be coded either by exonic

sequences (Cavaillé *et al*., 1996) or by intronic sequences (Tycowski *et al*., 1996); however, the latter is more commonly found in vertebrates (Kiss, 2002). For the PWS C/D-box snoRNAs, the function of only *SNORD115* has been identified in the regulation of alternative splicing for the serotonin receptor 2C mRNA (Kishore & Stamn, 2006), and the function of the other four classes of snoRNAs remains unknown. Therefore, understanding the evolutionary origins of PWS genes could provide significant insight for studies of gene function and imprinting mechanism. Previous studies suggested the evolutionary origin of *SNRPN* was ~95-125 mya (Rapkins *et al*., 2006). During evolution, *MKRN3* was derived from *MKRN1* by retrotransposition about 80-120 mya (Gray *et al*., 2000, 2001). *MAGEL2* and *NDN* were derived by retrotransposition from an ancestral X-chromosomal *MAGE* gene (Chai *et al*., 2001; Rapkin *et al*., 2006). The *C15orf2* gene was derived ~40 mya by retroviral insertion and is a primate-specific and intronless gene (Chai *et al*., 2001). Finally, the maternally expressed *UBE3A* and *ATP10C* genes were translocated to the current location (Chai *et al*., 2001). As a model system to understand the acquisition of new gene functions and the origin of genomic imprinting, we have focused on understanding the phylogenetic history of the key gene, *SNRPN*, associated with PWS (Nicholls & Knepper, 2001).

2.2 MATERIALS AND METHODS

2.2.1 Gene expression studies of 3 somatic cell hybrids

RNA was extracted from GM11715, A15, and A15-1 cell lines, and reversed transcribed into cDNA by Super Script®III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. The expression of PWS genes in GM11715, A15 and A15-1 cells were tested by primers listed in **Table 2.1** by regular PCR.

2.2.2 Chromatin immunoprecipitation (ChIP) in three somatic cell hybrids

Approximately 1×10^6 GM11715, A15, or A15-1 cells were plated on 35 cm^2 plates with regular media [Dulbecco's Modified Eagle Medium (DMEM) + 10% Fetal Bovine Serum (FBS) + 1% penicillin-streptomycin (P/S)]. When cells reached 70-80% confluence, formaldehyde was added to the regular media to make a final concentration of 1% and the plates were incubated at 37°C for 10 minutes to crosslink protein to DNA. After washing with 5 ml ice-cold PBS, another 5 ml ice-cold PBS was added and cells were scraped from the plates using a cell lifter, put into a 15 ml tube and sonicated to shear DNA. Samples were precleared at 4°C with protein G-agarose/salmon sperm beads for 1 h (Millipore). Protein-DNA was immunoprecipitated using the ChIP assay kit (Millipore) with anti-NRF1 (from Dr. Daniel Raines), anti-H3K4me1 [Histone H3 (mono methyl K4) antibody, Abcam, Catalog #: ab8895], anti-H3K4me3 [Anti- trimethyl-Histone H3 (Lys4), Millipore, Catalog #: 07-473], or anti-H3K9me2 [Anti-dimethyl- Histone H3 (Lys9), Millipore, Catalog #: 07-521]. Complexes were collected with Protein G agarose/salmon sperm beads and washed. Protein-DNA complexes were eluted off the beads and cross-links were reversed

by incubation at 65ºC overnight. On the next day, DNA was recovered by phenol-chloroform extraction and precipitated by ethanol. PCR was performed using the recovered ChIP DNA materials as templates and primers from **Table 2.2**. The PCR products were run on an agarose gel to examine the results.

Table 2.1. Primers for gene expression studies using somatic cell hybrids.

Primer name[a]	Primer number	Primer sequence
MKRN3 F	RN2998	5'-CACACGAGAAAGATATGGAAC-3'
MKRN3 R	RN2856	5′-GGAATGGTTGCAATTGGAAAG-3′
MKRN3-AS F	RN2976	5′-TCAAGTCCATCTTTGCTCCAC-3′
MKRN3-AS R	RN2977	5′-CTGTTGAGGCAGTGCTTCTG-3′
MAGEL2 F	RN2853	5′-GTGGAGGCACAGCCCTTGT-3′
MAGEL2 R	RN2854	5′-GGAACTGCACCAACGCATT-3′
NDN F	RN2851	5′-TGCTCCTGCAGAGTTTGGAA-3′
NDN R	RN2852	5′-GGCCTTGACTTTTCTTGGTTAGG-3′
MIRH1 F	RN2921	5′-ATATTGGGGCCTTCGATT-3′
MIRH1 R	RN2922	5′-ACTTGTCGAATTCCATCTGC-3′
PWRN1 F	RN2923	5′-TGTTTCCCAGACCGTAAC-3′
PWRN1 R	RN2924	5′-ACATCCAGTGGTAAGAAT-3′
C15orf2 F	RN2907	5′-CACCAGCAAGCCTATGAATTCC-3′
C15orf2 R	RN2925	5′-GTGGGGATGTGTAGACTGATGTC-3′
U1B F	RN3000	5'-AACTGTGGTCGCTGATCAATG-3'
U1B R	RN2999	5'-TCTGCGTTTGACTTGGACTTCC-3'
U1A F	RN2875	5′-GTCATTCTGCTTGCTGATCAAGA-3′
U1A R	RN2999	5'-TCTGCGTTTGACTTGGACTTCC-3'
SNRPN-SNURF exon 1-2 F	RN3001	5'-TGACGCATCTGTCTGAGGAG-3'
SNRPN-SNURF exon 1-2 R	RN2999	5'-TCTGCGTTTGACTTGGACTTCC-3'
SNORD107 F	RN1455	5'-GGCTAGGTTCATGATGACAC-3'
SNORD107 R	RN1456	5'-GGCACACTGACTGGATTTCA-3'
SNORD64 F	RN1433	5'-GGATTTGTGATGAGCTGTGT-3'
SNORD64 R	RN1434	5'-GGACTTCAGAGTAATCACGT-3'
SNORD109 F	RN1583	5'-GGATCGATGATGAGAATAA-3'
SNORD109 R	RN1584	5'-GAACCTCAGATTGACATGT-3'
SNORD116 class I F	RN1579	5'-ATGATGATGAGTCCCC-3'
SNORD116 class I R	RN1580	5'-GACCTCAGTTCCGATGAG-3'
SNORD116 class II F	RN1581	5'-GATCGATGATGACTTCCA-3'
SNORD116 class II R	RN1582	5'-GGACCTCAGTTCGACGAG-3'
IPW F	RN3002	5'-ATGACTTCCTGGGAACTCTTC-3'
IPW R	RN2874	5′-TGGCACCAACTCAACAAATCC-3′
SNRPN-SNURF exon 63-65 F	RN2970	5′-CTGAAGCTCAGGCCATTCCT-3′
SNRPN-SNURF exon 63-65 R	RN2971	5′-GTCTTCCTCCAGGCTCACTG-3′
SNORD115 F	RN1435	5′-GGGTCRATGATGAGAACCTT-3′
SNORD115 R	RN1436	5′-GGGCCTCAGCGTAATCCTAT-3′
SNRPN-SNURF exon 142 F	RN2972	5′-CACAGCTGACACACCCAGATATC-3′
SNRPN-SNURF exon 143 R	RN2973	5′-CCCTGAAGTTTCCTTGAAGTTGTT-3′
SNRPN-SNURF exon 144 R	RN3069	5′-TGAAGTGGAGGATCAGATTCCAG-3′
SNRPN-SNURF exon 145 R	RN3070	5′-GTTACTTAATCATACAGTAAGCTG-3′

Table 2.1. (continued)

Primer name[a]	Primer number	Primer sequence
SNRPN-SNURF exon 147-148 F	RN2974	5′-AGAAAAGGCGCAATGAAAGA-3′
SNRPN-SNURF exon 147-148 R	RN2975	5′-GGAGAAGGGCCATAGACTCC-3′
UBE3A F	RN2980	5'-AACAAGAAAGGCGCTAGAATTG-3'
UBE3A R	RN2981	5'-TAACTTTCCGGAAGCTCTGTAC-3'
ATP10A F	RN2978	5'-CCTGGCTCAACTGGATAACG-3'
ATP10A R	RN2979	5'-AACTGACGTGCCAGCTGAAG-3'
5.8S rRNA F	RN2849	5'-CGACTCTTAGCGGTGGATCA-3'
5.8S rRNA R	RN2850	5'-GACGCTCAGACAGGCGTAG-3'

[a] Abbreviations: F: forward PCR primer; R: reverse PCR primer. Gene sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/Homo_sapiens/Info/Index).

Table 2.2. Primers for ChIP studies using somatic cell hybrids.

Primer names	Primer number	Primer sequence
SNURF-SNRPN enhancer F	RN2683	5′-CAGGTCTTGGAAGGCTATGTCG-3′
SNURF-SNRPN enhancer R	RN2684	5′-CTCCCCACTGGCGGCTCTAC-3′
SNURF-SNRPN promoter F	RN2877	5′-TGCACTGCGGCAAACAAG-3′
SNURF-SNRPN promoter R	RN3059	5′-AGACAGATGCGTCAGG-3′
U1A promoter F	RN2967	5′-CTGAAGCCTTTGGGATTTTCTA-3′
U1A promoter R	RN2968	5′-AAACAGCTCCAGAGGGAAGAC-3′
U1A (upstream *U1A* promoter) F	RN3266	5′-CGCCATAACCGGAAGAGATC-3′
U1A (upstream *U1A* promoter) R	RN3265	5′-CCCTTTCCCCCTGCTATTTC-3′
U1B (upstream *U1B* promoter) F	RN3263	5′-TGGTCTCCCGTCACCGTTT-3′
U1B (upstream *U1B* promoter) R	RN3264	5′-ACTCCTAGCAGACGCCATAAGC-3′
NDN promoter F	RN2894	5′-CCTAGATCTTCTCAGCCCAAACA-3′
NDN promoter R	RN2895	5′-TCCATGGCGAGGCTTCAC-3′
CpG islet (*MKRN3*) F	RN3033	5′-ATTAAACGCGAGTGTCCAGAAT-3′
CpG islet (*MKRN3*) R	RN3034	5′-CAAACCTGTGAACTAGTCGGTGTA-3′
NRF1 cluster	RN3006	5′-CCTTCTGGCGCATGCGCAGTCTCT-3′
NRF1 cluster	RN2960	5′-AGTGGACTCTGCAATGCTGTAA-3′
MKRN3 promoter F	RN3053	5′-GAAATCGTGTGAGAAGGGACTTAG-3′
MKRN3 promoter R	RN3054	5′-CCTCTGACTGTGTGTTCCTACCTA-3′
MIRH1 promoter F	RN3040	5′-CGTGAGTTAAGAGTGTGAAGGAGA-3′
MIRH1 promoter R	RN3041	5′-CAAACAGTCTGCTCTGAGTCTTGT-3′
PWRN1 promoter F	RN3581	5′-GGCGGTGAGATCTACAGGAG-3′
PWRN1 promoter R	RN3582	5′-GTCTGGGAAACACGGAGAAATC-3′
MAGEL2 promoter F	RN3579	5′-CAGCCTCTGATCCTGCAAATC-3′
MAGEL2 promoter R	RN3580	5′-GTCTGCCAAGTCAGGGGAGTG-3′

[a] Abbreviations: F: forward PCR primer; R: reverse PCR primer. Gene sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/Homo_sapiens/Info/Index).

2.2.3 Bioinformatics and phylogenetic analyses

Gene sequences were obtained from the Ensembl genome browser and putative NRF1 binding sites were localized within potential gene regulatory regions by examining 5 kb upstream of

the TSS, exon 1, and all introns and 5 kb downstream of 3' UTR of the gene. In this second step, sites matching a GCGCAYGCGC motif, allowing at most one mismatch in one "GCGC" and/or one mismatch in the "AY" segment (on either strand), were selected as potential NRF1 motifs. Second, the sequence spanning the NRF1 motif(s) was used for BLAST (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) searches in the non-redundant (NR) and whole genome shotgun (WGS) databases to determine conservation of NRF1 motif(s) across mammalian sequences. Sequences were aligned by ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html) and conserved motifs were highlighted and analyzed by STAMP or other bioinformatic tools.

Genomic sequences for the *M. domestica SNRPN* and *SNRPB'* loci were obtained by NCBI BLAST with the *M. domestica SNRPB'* cDNA sequence (Gray *et al*., 1999b). Sequences spanning each gene were exon-masked and/or Repeatmasked [using the "mammal (other than below)" in "DNA source" option] and used for BLAST. *M. domestica*-specific repetitive sequences were then removed *in silico* when necessary. Candidate non-coding conserved sequences (NCCS) were subsequently analyzed further by BLAST and visual inspection. *SNORD119* snoRNA gene sequences from different species were also obtained from NCBI BLAST searches. *SNORD64*, *SNORD107*, *SNORD109*, *SNORD115* and *SNORD116* human copies were obtained from the NCBI gene database, and other eutherian copies of *SNORD64*, *SNORD107* and *SNORD109* were obtained by BLAST searches in the WGS database. Names of different classes of PWS snoRNAs were used according to official nomenclature from PubMed. *M. domestica SNRPB'* exons and introns sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/index.html). All sequence was transformed into FASTA format using ReadSeq (http://searchlauncher.bcm.tmc.edu/seq-util/readseq.html) and alignments were performed using ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html). The alignment of *SNORD116* was

derived from both ClusterW and Runte *et al*. (2001). Alignments were adjusted as needed to maximize parsimony.

2.2.4 Calculation of minimum free energy in prediction of RNA-RNA interaction

The minimum free energy of interaction for the snoRNA and predicted target RNA was calculated by RNAhybrid (Rehmsmeier *et al*., 2004), which is based on the classical RNA secondary structure prediction algorithm to two sequence (Zuker &Stiegler, 1981) and energy parameters of RNA-RNA interaction by Mathews *et al*. (1999). SnoRNAs and predicted 28S rRNA sequence were submitted in RNAhybrid (http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/submission.html) to predict the minimum free energy.

2.2.5 Sequence cloning and analyses

PCR primers (**Table 2.3**) flanking *Didelphis virginiana SNORD119N* and *SNORD119B* were used to clone the respective fragments from *D. virginiana* genomic DNA. The two PCR fragments were then subcloned in to a TOPO TA vector (Invitrogen) for sequencing according to the manufacturer's instructions.

2.2.6 Marsupial, mouse and human RNA and DNA studies

RNA was extracted from *M. domestica* tissue samples with Trizol (Invitrogen) followed by DNase I treatment (Promega). The latter step to remove DNA was critical for studies of small RNA expression where no introns were present. A mouse total RNA panel (10 tissues) and FirstChoice™

Human total RNA survey panel (20 tissues) were purchased from Ambion. PCR and quantitative (Q)RT-PCR primers (**Table 2.3**) were designed by the online primer design program Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and checked by NetPrimer (http://www.premierbiosoft.com/netprimer/index.html). 1 μg of each human, mouse, and *M. domestica* RNA sample was reverse-transcribed (SuperScript II, Invitrogen) into cDNA using hexamer oligonucleotides or specific snoRNA reverse primers in a 20 μl reaction. cDNA was amplified by *Taq* DNA polymerase (Promega), using 94°C for 5 minute, 35 cycles of 94°C for 30 seconds, 50°C (snoRNA primers)/53°C (5.8S rRNA primers)/55°C (*SNRPN*/*SNRPB'* PCR primers) for 30 seconds, and 72°C for 1 minute, then followed by a 10-min final extension at 72 °C. PCR products were analyzed by agarose gel electrophoresis.

2.2.7 Verification of *SNORD119B* and *SNORD119N* snoRNA specific RT-PCR assays by restriction enzyme fragment length variants (RFLV)

The *SNORD119B* and *SNORD119N* snoRNA gene sequences were aligned by Clustal W. The *Hae*III enzyme was chosen to differentiate *SNORD119B* snoRNA from *SNORD119N* snoRNA RT-PCR products. The enzyme digested RT-PCR products were analyzed by agarose gel electrophoresis.

Table 2.3. Primers for evolutionary studies of the PWS SNRPN locus.

Species, target gene, and orientation[a]	Primer	Primer Sequence
Human *SNORD119* snoRNA F	RN2731	5′-GCTGGATTAATGATGAGAT-3′
Mouse *Snord119* snoRNA F	RN2733	5′-GCTGGATTCATGATGAAAT-3′
Human/Mouse *SNORD119* snoRNA R	RN2732	5′-ATCTCAGAGTAATCCTGC-3′
M. domestica SNORD119N snoRNA F	RN2734	5′-GCTGGATTGGTGATGAAACC-3′
M. domestica SNORD119B snoRNA F	RN2736	5′-GCTGGATTAGTGATGAAACA-3′
M. domestica SNORD119N/B' snoRNA R	RN2735	5′-ACCTCAGAGTAATCCTGCT-3′
M. domestica SNRPN (exon 5) F	RN2787	5′-CACAGCCAGTATCGCAGGAGC-3′
M. domestica SNRPN (exon 7) R	RN2788	5′-CAGAGGCAGCATCACAGCAAA-3′
M. domestica s *SNRPB'* (exon 5) F	RN2789	5′-CCCAGGGAAGAGGAACTGT-3′
M. domestica SNRPB' (exon 7) R	RN2790	5′-GCAAGCATCTGAGCAAAGC-3′
Human *SNRPB/B'* (exon 6) F	RN2791	5′-CTGGTATGAGACCTCCTATGG-3′
Human *SNRPB/B'* (exon 7) R	RN2792	5′-CCACAAGGAGATAAAAGGACT-3′
Mouse *Snrpb/b'* (exon 6) F	RN2793	5′-CTGGCCGAGGAACTCCAAT-3′
Mouse *Snrpb/b'* (exon 7) R	RN2794	5′-GGAAGGAAACAGGCAAGGA-3′
Human/Mouse/ *M. domestica* conserved 5.8S rRNA F	RN2849	5′-CGACTCTTAGCGGTGGATCA-3′
Human/Mouse/ *M. domestica* conserved 5.8S rRNA R	RN2850	5′-GACGCTCAGACAGGCGTAG-3′
M. domestica GAPDH cDNA primer (QPCR) F	RN3369	5′-GCCGAGTACGTTGTGGAGTCC-3′
M. domestica GAPDH cDNA primer (QPCR) R	RN3370	5′-AGGGGGCAGAGATAATGACG-3′
M. domestica SNRPN (exon 5 QPCR) F	RN3267	5′-AGCCAGTATCGCAGGAGCC-3′
M. domestica SNRPN (exon 5 QPCR) R	RN3268	5′-GGTGTCCCTCTGCTCACGG-3′
M. domestica SNRPB' (exon 5 QPCR) R	RN3269	5′-AGCCAGCATTGCAGGGGCTC-3′
M. domestica SNRPB' (exon 5 QPCR) R	RN3270	5′-GGTGCCCCTCGACCCATTG-3′
M. domestica SNORD119N (QPCR) F	RN3365	5′-GCTGGATTGGTGATGAAACCTA-3′
M. domestica SNORD119N (QPCR) R	RN3366	5′-CCTCAGAGTAATCCTGCTGAG-3′
M. domestica SNORD119B (QPCR) F	RN3367	5′-GCTGGATTAGTGATGAAACCTG-3′
M. domestica SNORD119B (QPCR) R	RN3368	5′-ACCTCAGAGTAATCCTGCTAAC-3′
D. virginiana *SNRPN* (exon 5) F	RN3362	5′-GCGGTTGCTGCCACAGCCAG-3′
D. virginiana SNRPN (exon 6) R	RN3371	5'-TCCCCATTGGTGCCCCTCGAC-3'
D. virginiana SNRPB' (exon 5) F	RN3364	5′-GCCGCCACAGCCAGCATTGC-3′
D. virginiana SNRPB' (exon 6) R	RN3363	5′-AGGGGGTCTCATGCCAGGAGG-3′

[a] Abbreviations: F: forward PCR primer; R: reverse PCR primer. Gene sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/Homo_sapiens/Info/Index).

2.2.8 Quantitative analysis of mRNA expression

1 μg of each *M. domestica* RNA sample was reverse transcribed (SuperScript III, Invitrogen) into cDNA using hexamer oligonucleotides and oligo (dT) primers in a 20 μl reaction. Expression of *M. domestica SNRPN*, *SNRPB'*, *SNORD119N*, *SNORD119B* genes using respective primers (**Table 2.3**) was analyzed by quantitative real-time PCR (QPCR) in Applied Biosystems 7300 Real Time PCR system. SYBR green (Power SYBR® Green PCR Master Mix, Applied Biosystems) was used as a fluorescent dye and each cDNA sample was run in triplicate to calculate the statistical significance of the data. Threshold cycle (C_T) was reported from the Applied Biosystems 7300 Real Time PCR system (**Table 2.4** & **2.5**). Outliers in the C_T dataset were determined by Analyze-it for Microsoft Excel 2003 to generate an outlier box plot and the observations outside 1.5 interquartile ranges (IQRs) were defined as outliers. In each tissue sample, the ΔC_T value was obtained by normalizing each gene of interest against *M. domestica GAPDH* data for each assay. The $\Delta\Delta C_T$ value was obtained by subtracting the ΔC_T values for each sample from the ΔC_T values of *GAPDH* in each corresponding assay to calculate a relative $\Delta\Delta C_T$. All $\Delta\Delta C_T$ values were converted to fold change. 2-sample independent t-test was used in Microsoft Excel 2003 to compare the expression difference ($p<0.05$) between 1) *SNRPN* and *SNRPB'*; 2) *SNORD119N* and *SNORD119B* in each tissue sample. Comparisons of gene expression measured by QRT-PCR in brain and non-brain tissues between 1) *SNRPN* and *SNRPB'*; 2) *SNORD119N* and *SNORD119B* were made using 2-way ANOVA (SPSS v13.0) with repeated measures.

Table 2.4. Average C_T for 5 genes in brain tissues of *M. domestica*[a].

C_T	11wk ♀ brain			11wk ♂ brain				Adult ♂ brain			
	Exp 1	Exp 2	Exp 3	Exp 1	Exp 2	Exp 3	Exp 4	Exp 1	Exp 2	Exp 3	Exp 4
SNRPN	27.64	27.39	28.47	28.83	27.18	28.86	27.15	28.48	27.98	28.90	28.67
SNRPB'	29.00	28.59	29.82	30.23	28.02	29.47	27.88	30.57	29.46	30.72	30.20
SNORD119N	31.67	32.21	33.14	32.39	31.30	33.01	31.75	33.26	32.91	33.07	28.76[b]
SNORD119B	33.66[b]	35.60	35.37	31.62[a]	32.85	33.26	32.00	33.12[a]	34.13	33.24	32.80
GAPDH	22.83	23.19	24.32	24.38	23.72	24.19	22.77	24.40	24.22	25.11	24.42

[a] Abbreviation and symbols: C_T: threshold cycle; wk: week; ♀: female; ♂: male; exp: experiment.

[b] These 3 *SNORD119B* assays were ran separately with *GAPDH* control average C_T values of 22.94, 24.67, 24.75 for 11wk ♀ brain, 11wk ♂ brain, Adult ♂ brain, respectively. [b] The results from this assay were defined as outliers by Analyze-it for Microsoft Excel and not included in the data Figure 2.13.

Table 2.5. Average C_T for 5 genes in non-brain tissues of *M. domestica*[a].

C_T	Liver			Spleen			Kidney			Muscle		
	Exp 1	Exp 2	Exp 3	Exp 1	Exp 2	Exp 3	Exp 1	Exp 2	Exp 3	Exp 1	Exp 2	Exp 3
SNRPN	25.35	24.72	24.84	25.23	25.45	25.86	29.41	28.52	28.71	25.27	27.44	24.88
SNRPB'	24.44	23.68	23.63	23.09	23.08	23.24	28.78	27.53	27.63	25.19	27.28	24.83
SNORD119N	29.95	29.22	29.06	28.38	28.48	28.66	32.00	29.56	31.16	31.93	38.38	33.08
SNORD119B	32.42	31.65	31.62	29.16	29.52	29.43	33.52	33.00	32.54	35.13	38.29	38.93
GAPDH	19.73	19.18	18.92	21.08	21.78	21.76	24.41	23.42	22.96	16.96	20.26	17.06

[a] Abbreviation and symbols: C_T: threshold cycle; exp: experiment.

2.3 RESULTS

2.3.1 Expression profile of PWS genes in three somatic cell hybrids

Expression studies showed that most PWS genes are expressed robustly in GM11715 cells (**Figure 2.3**). In contrast, the homologous mouse genes are not expressed (unpublished data). For A15 cells, previous studies have shown high expression of key PWS genes, including *SNRPN* and *NDN* (Gabriel *et al*., 1998).

2.3.2 NRF1 ChIP results and chromatin status of the 2 Mb PWS region in somatic cell hybrids

Our laboratory has previously demonstrated multiple functional NRF1 sites in the 2 Mb PWS region (Stefan *et al*., in preparation). In this study, I verified the NRF1 regulatory function on PWS genes in a mouse/human somatic cell hybrid cell line, GM11715, by performing NRF1 ChIP assay. As expected, since the *SNURF-SNRPN*-snoRNA polycistronic locus is well expressed in these cells, I found by ChIP assays that NRF1 binds strongly to the *SNURF-SNRPN* promoter and enhancer (**Figure 2.4**). NRF1 also bound to regulatory elements associated with other expressed genes in these cells (**Figure 2.4**), including the *NDN* promoter, the *U1A* promoter and strongly to an upstream putative enhancer, and the *PWRN1* promoter (the first such confirmation), but only weakly to the *MKRN3* enhancer (this Is not surprising since this gene is not expressed in these cells). I also tested the *MAGEL2* promoter, a gene expressed robustly in GM11715 cells, as there

was a weak sequence match to the NRF1 motif but NRF1 does not bind at this element, a result that further helps define functional NRF1-binding motifs. Consistent with no NRF1 binding site in the *MIRH1* promoter, there is no binding of NRF1 (**Figure 2.4**). Surprisingly, since *MIRH1* is expressed at good levels in GM11715 cells, the "NRF1-cluster" (a strong *Mirh1* enhancer *in vitro*) is only very weakly bound by NRF1 despite strong binding of NRF1 to the "NRF1 cluster" in SK-N-SH neuroblastoma cells. Histone code studies have indicated that active chromatin is associated with H3K4me1 (Heinzman *et al*., 2007), and using GM11715 cells I found that all the gene promoters and enhancers in the PWS domain have an active chromatin status [**Figure 2.5(a)**]. Additionally, positive ChIP results using H3K4me3 and negative results with H3K9me2 further support the active chromatin status in this region [**Figure 2.5(b) & (c)]**. Nevertheless, the "NRF1 cluster" and *MKRN3* regulatory regions are only weakly positive. Therefore, we concluded that there are two domains in the 2 Mb PWS imprinted region: one is the *U1A*-*SNURF*-*SNRPN*-snoRNA segment which is highly expressed and strongly binds NRF1 to all promoter and enhancer elements in GM11715 cells, and a more proximal domain containing genes expressed at lower levels (e.g., *NDN*, *MIRH1*) or not at all (e.g., *MKRN3*) with low levels of NRF1 binding (e.g., *NDN* promoter, *MKRN3* enhancer, "NRF1 cluster"). ChIP assays were also performed on both A15 and A15-1 cells to test the binding of NRF1 to NRF1 sites on the human chromosome 15 of PWS region. Unfortunately, our data showed that NRF1 also only bound weakly to the "NRF1 cluster" in the PWS region of A15 or A15-1 [**Figure 2.6(a) & (b)**].

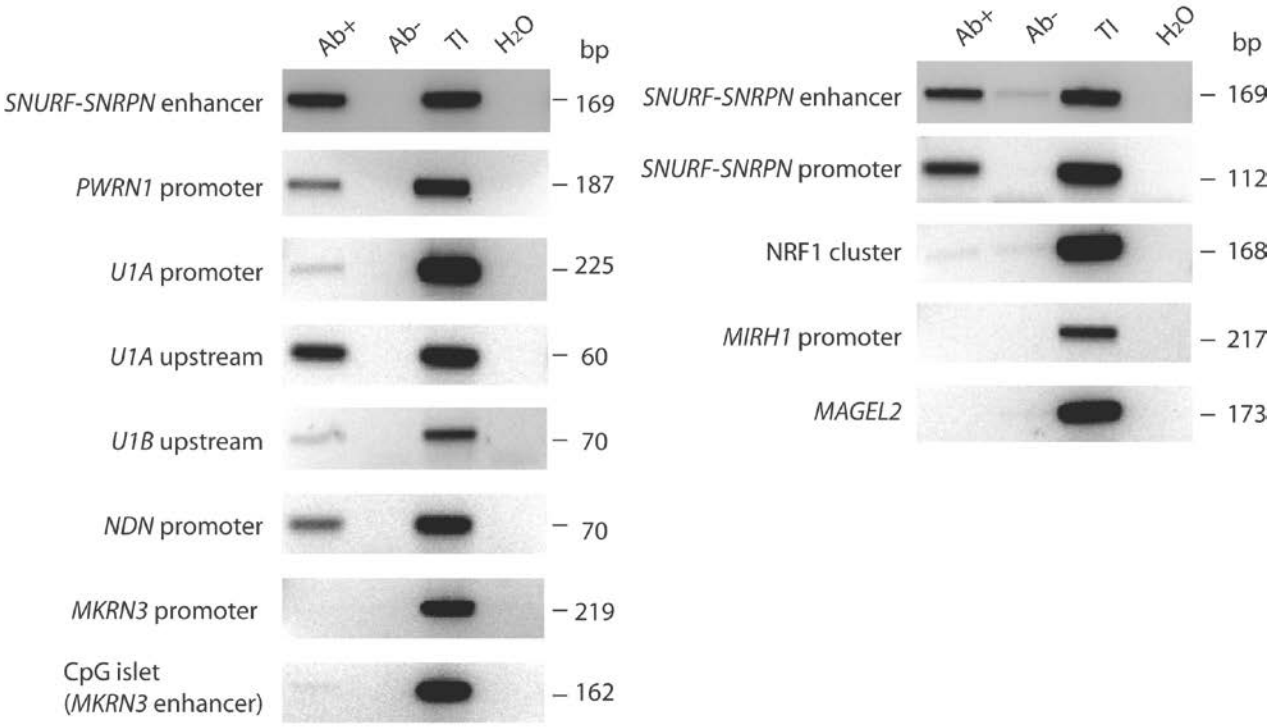


Figure 2.4. NRF1 ChIP assay using GM11715 cells. Abbreviations: Ab+: antibody positive; Ab-: antibody negative; TI: total input; bp: base pair.

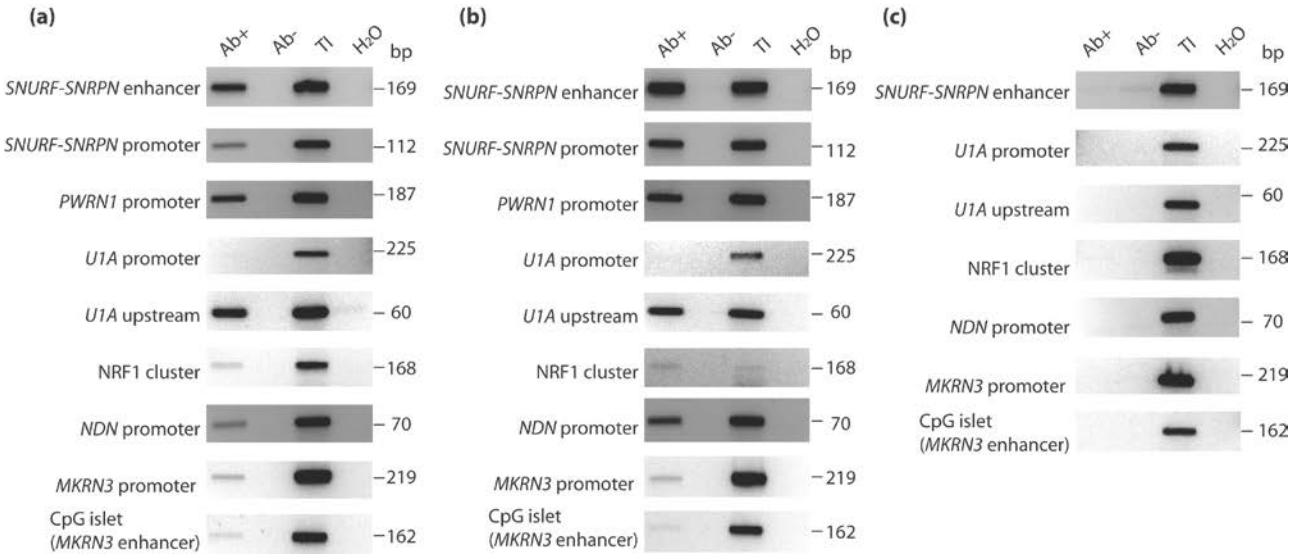


Figure 2.5. ChIP assays using GM11715 cells for (a) H3K4me1, (b) H3K4me3, and (c) H3K9me2 antibodies. Abbreviations: Ab+: antibody positive; Ab-: antibody negative; TI: total input; bp: base pair.

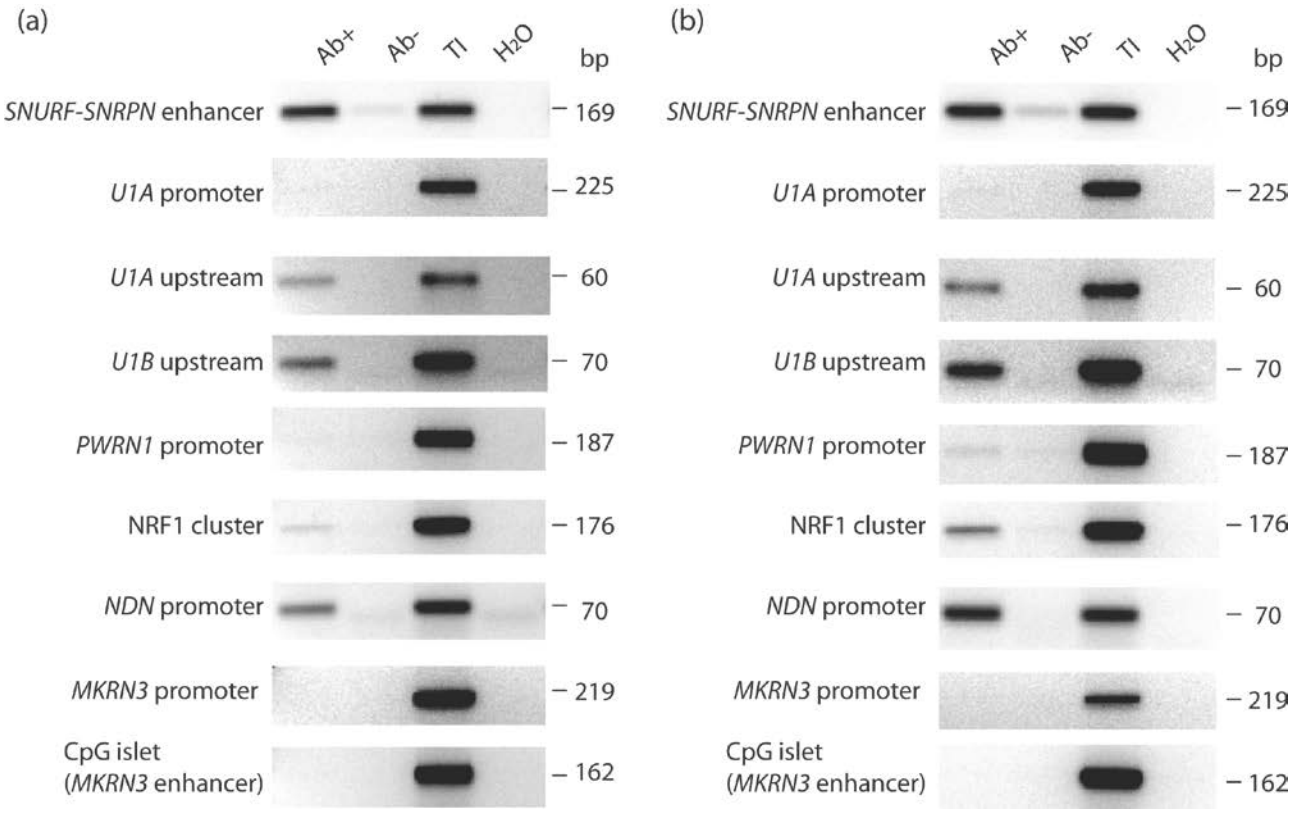


Figure 2.6. NRF1 ChIP assays using (a) A15 and (b) A15-1 cell lines. Abbreviations: Ab+: antibody positive; Ab-: antibody negative; TI: total input; bp: base pair.

2.3.3 Evolutionary studies of the PWS *SNRPN* locus and snoRNAs

2.3.3.1 Intragenic snoRNAs in the duplicated *SNRPB'* and *SNRPN* loci of marsupials

In *M. domestica*, *SNRPN* and *SNRPB'* are adjacent paralogs on chromosome 1 (Rapkins *et al*., 2006). By sequence analysis across species, we found three conserved elements, which were in intron 5 of *SNRPN*, exon 2b of *SNRPB'*, and intron 5 of *SNRPB*' (**Figure 2.7**), respectively. Further analysis confirmed the conserved sequences in intron 5 of *SNRPN* and intron 5 of *SNRPB*' were paralogous genes, both of which were C/D box snoRNAs, and their ortholog in human was named *SNORD119* (Yang *et al*., 2006). Because *SNRPN* and *SNRPB'* were paralogous genes, we named the intron 5 snoRNA in *SNRPN* as *SNORD119N* and the intron 5 snoRNA in *SNRPB'* as *SNORD119B*. To further examine these results in other marsupial species, we verified the existence of the snoRNAs in *Didelphis virginiana* (Virginia opossum) by PCR cloning and in *Macropus eugenii* (tammar wallaby) by sequence BLAST (**Figure 2.8**).

2.3.3.2 Molecular evolution of the *SNORD119* snoRNA within *SNRPB'* intron 5 in vertebrates

Since the protein of *SNRPB'*, SmB', is an indispensable protein that constitutes one of the major proteins in the assembly of the Sm spliceosome (Matera *et al*., 2007) and snoRNAs have important functions in rRNA modification (Maden, 2001), it was not surprising that we found *SNRPB'* conserved across species, from yeast to eutherian mammals. *SNORD119B* orthologs were found consistently in intron 5 of the *SNRPB'* gene of all analyzed eutherian mammals, reptiles, *Xenopus tropicalis*, and zebrafish, in intron 6 of the *SNRPB'* gene of Atlantic salmon, but not in any intron of birds [**Figure 2.9(a)**].

By sequence BLAST, we determined that genes (*CD20-ELMO2*) upstream of *M. domestica SNRPN* were syntenic to human chromosome 20q13.12 and to mouse 2H3, while those downstream

of *M. domestica SNRPN* (*SNRPB'-STK35*) were syntenic to human chromosome 20p13 and to mouse 2F3, and the block of *M. domestica SNRPN* was syntenic to the PWS region of human chromosome 15q11.2 and to mouse 7C (**Table 2.6**). We found large blocks of repetitive elements on the 5' upstream (~47k, 96% repetitive) and 3' downstream (~6k, 67% repetitive) sides of *M. domestica SNRPN*, respectively (**Table 2.6**).

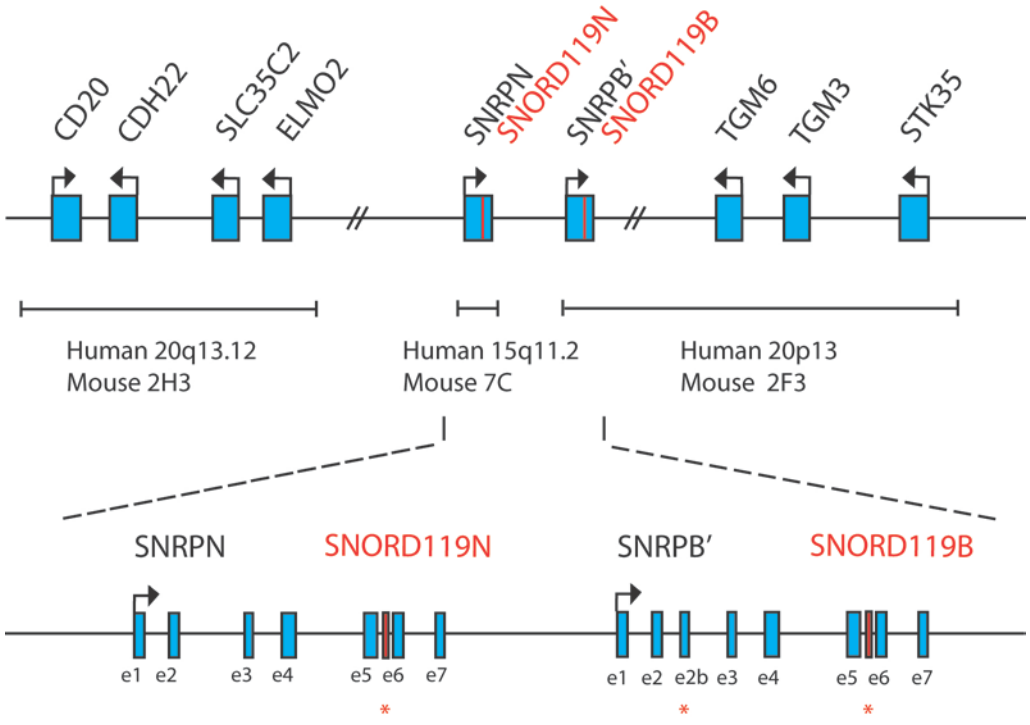


Figure 2.7. Genetic maps of the *SNRPN--SNORD119N* and *SNRPB'--SNORD119B* loci, and syntenic sequences in human and mouse. Arrows indicated transcriptional direction. Red asterisks indicated conserved regions.

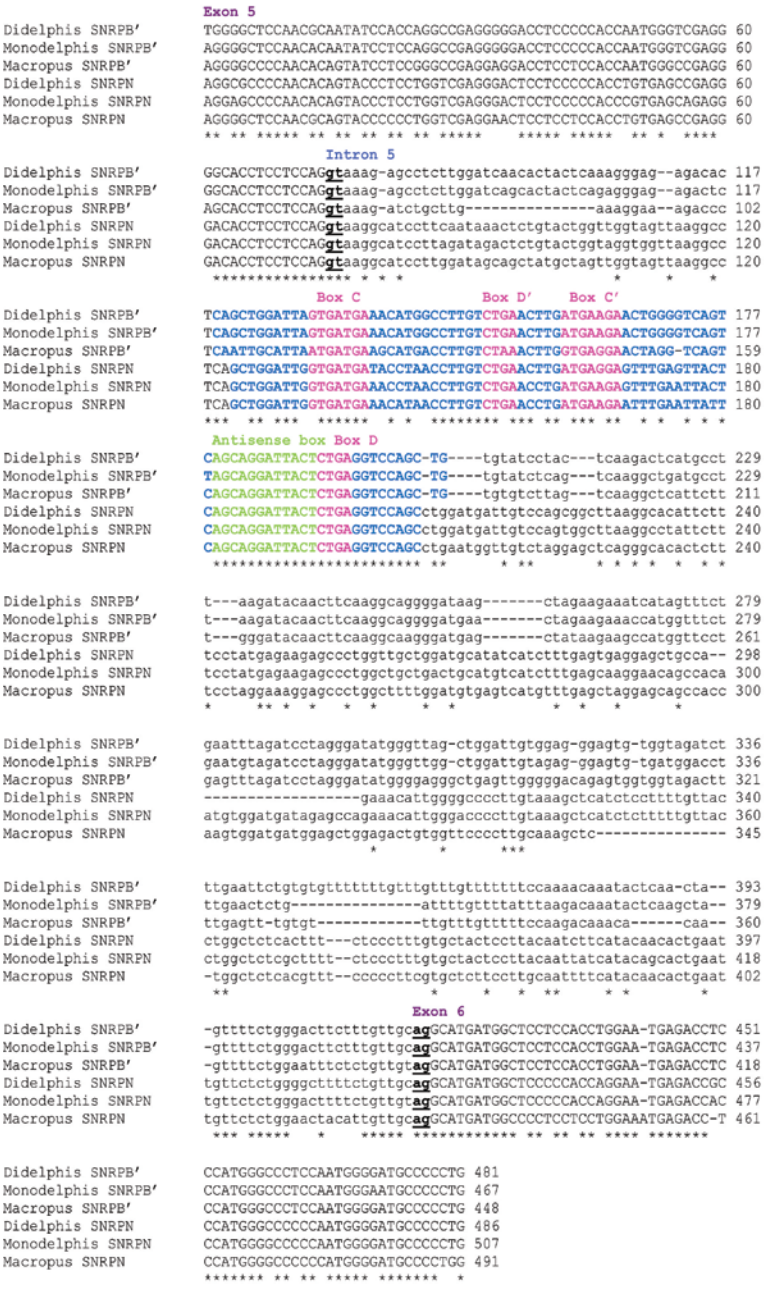


Figure 2.8. DNA sequence comparison of *SNORD119B* and *SNORD119N* in three species of metatherians, *M. domestica*, *D. virginiana*, and *M. eugenii*. Intron sequence is in small case, with exon-intron and intron-exon boundaries were in bold black and underlined. The *SNORD119N* and *SNORD119B* alignment is in bold blue. Box C/C' and box D/D' are in bold pink and the antisense box is in bold green.

Table 2.6. Analysis of the *ELMO2-SNRPN-SNRPB'-TGM6* genomic interval in *M. domestica*.

Region	Extracted feature[a]	Size (bp)[a]	Repetitive	Unique	Features identified
ELMO2-SNRPN	None	47,876	96%	4%	Gene promoters & FLJ13305 pseudogene[b]
SNRPN intragenic	Exons, snoRNA	12,976	44%	56%	*TAF9* pseudogene[c]
SNRPN-SNRPB'	None	6,210	67%	33%	None
SNRPB' intragenic	Exons, snoRNA	10,530	36%	64%	None
3' of *SNRPB'*	None	28,470	69%	31%	3' end of *TGM6* (exon 9-12)

[a] Size of region analyzed after extracting (masking) exons and snoRNAs. For this analysis, we masked *SNRPN* exons 1-7, *SNRPB'* exons 1-7, *SNRPB'* exon 2b, *SNORD119N* and *SNORD119B* sequences.

[b] The *SNRPN* and *ELMO2* promoters have only 363-bp and 659-bp unique sequence, respectively. The *ELMO2* promoter has two elements conserved in mammals. The 1.55 kb FLJ13305 pseudogene is located 36.2-kb 5' of *SNRPN*.

[c] Located in *SNRPN* intron 4, on the complementary strand.

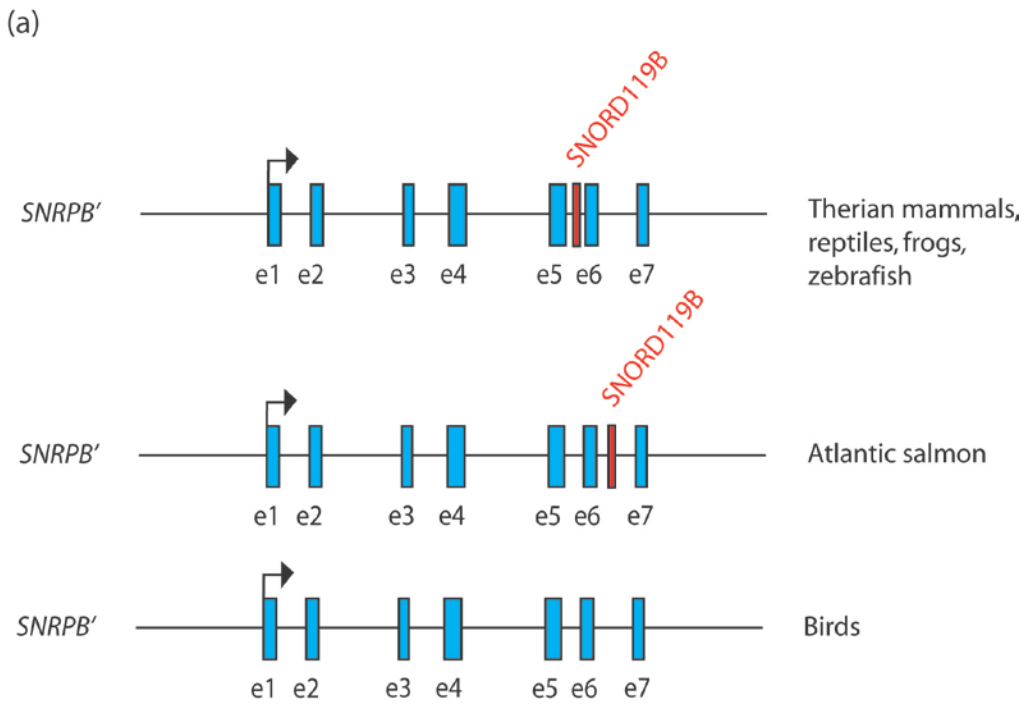


Figure 2.9. (continued below)

(b)

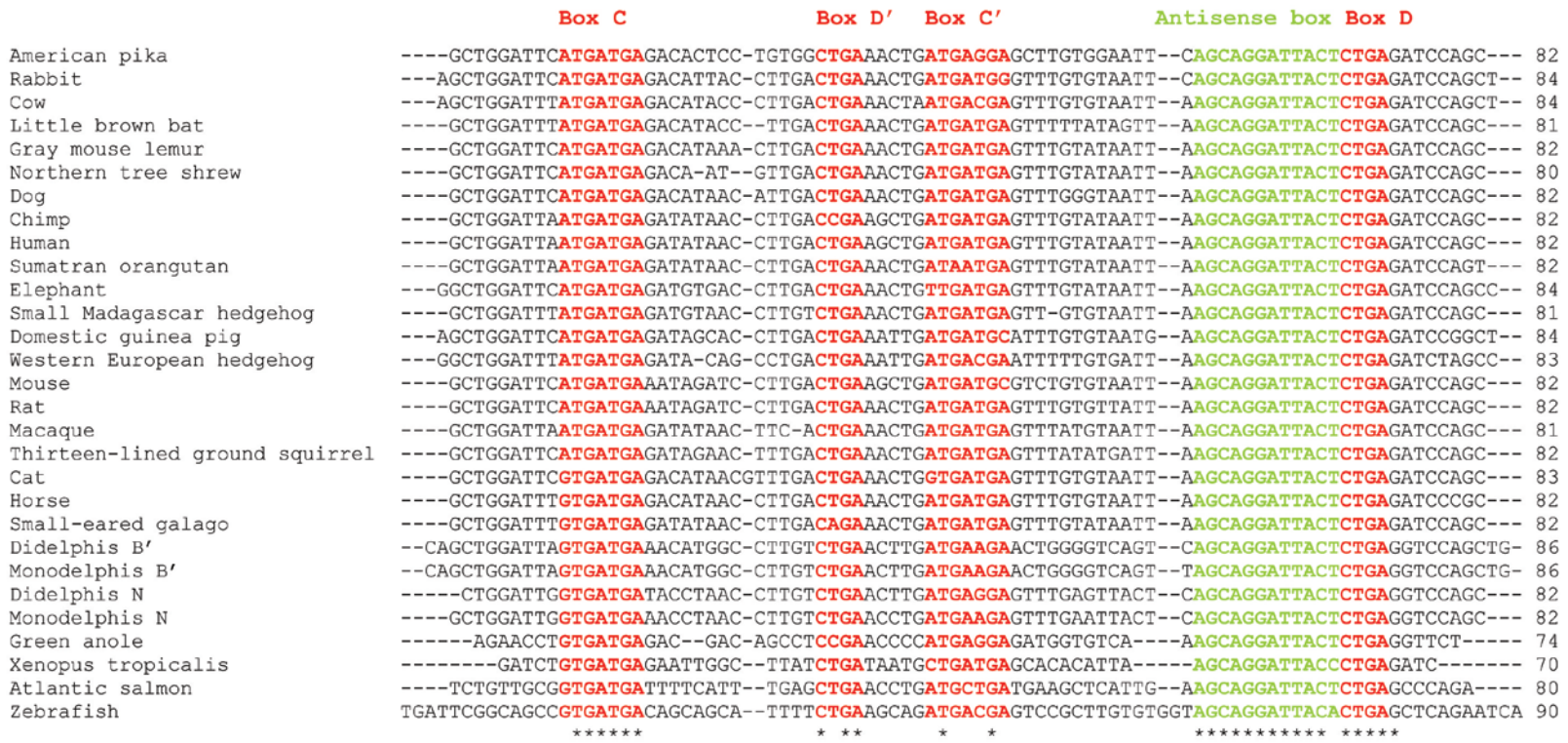


Figure 2.9. Phylogenetic analyses of the *SNORD119B* snoRNA. (a) Cartoon of *SNRPB'* gene structure of therian mammals, reptiles, an amphibian and zebrafish with the *SNORD119* snoRNA in intron 5, whereas the *SNORD119* snoRNA is in intron 6 of Atlantic salmon. In contrast, the *SNRPB'* gene in birds has no intragenic snoRNA. (b) Clustal W multisequence alignment of *SNORD119* snoRNAs. Box C/C' and box D/D' are highlighted in bold red: from 5' to 3', these are box C, box D', box C' and box D, respectively. The antisense box is highlighted in bold green and is 5' of box D.

2.3.3.1 Expression studies of *SNORD119* orthologs and paralogs

Expression studies by RT-PCR showed that *SNORD119* snoRNA was ubiquitously expressed in 20 analyzed human tissues [**Figure 2.10 (a)**] and 10 analyzed mouse tissues [**Figure 2.10(b)**]. In *M. domestica*, *SNORD119N* and *SNORD119B* were ubiquitously expressed in 8 tissues from one female animal and 4 brain tissue samples from 4 animals (11 weeks postpartum male and female, adult male and female, respectively) [**Figure 2.11(a) & (b)**]. Because *SNORD119N* and *SNORD119B* were highly conserved sequences (**Figure 2.8**) and the PCR primers for *SNORD119N* and *SNORD119B* only differ in two non-contiguous nucleotides, to confirm we had tested specific gene expression, we used *Hae*III enzyme digestion on both PCR products and verified the primers were specific to *SNORD119N* and *SNORD119B* PCR products [**Figure 2.12(b)**] in *M. domestica*.

By quantitative real-time PCR (QRT-PCR) studies of gene expression in *M. domestica*, our results showed that in 3 brain tissues, *SNRPN* was expressed at a higher level than *SNRPB'* (11 weeks postpartum female, $P<0.05$, and adult male, $P<0.001$, but not in 11 weeks postpartum male, $P=0.116$), while in the 3 non-brain tissues, liver, kidney and spleen, *SNRPB'* expression was significantly higher than *SNRPN* [liver, spleen, and kidney, all $P<0.05$; **Figure 2.13(a)**]. These results were consistent with protein studies of SmN and SmB in mice (Gray *et al*., 1999b). In addition, in 3 of 4 female tissues (one juvenile brain and all non-brain tissues were from the same adult animal) analyzed, *SNORD119N* gene expression was significantly higher than *SNORD119B* (11 weeks postpartum female, $P<0.001$; liver, $P<0.0001$; spleen, $P<0.05$; but in kidney, $P=0.141$), except in two male brain samples [11 weeks postpartum male, $P=0.990$, and adult male, $P=0.808$, **Figure 2.13(b)**]. This suggests a different expression level of *SNORD119N* and *SNORD119B* in male versus female in *M. domestica*; however, larger numbers of animals would be needed to

further examine this possibility. By 2-way ANOVA analyses to compare gene expression in brain tissues, *SNRPN* gene expression was significantly higher than *SNRPB'* [F(2, 1)=31.10, $P<0.0001$), but *SNORD119N* and *SNORD119B* gene expression was not [F(2, 1)=0.71, $P=0.414$) due to the lack of expression difference in the two male brain tissues [**Figure 2.13(b)**]. Similarly, 2-way ANOVA analyses comparing gene expression in non-brain tissues (liver, spleen and kidney) showed that *SNRPN* expressed significantly lower than *SNRPB* [F(2, 1)=45.08, $P<0.0001$] and *SNORD119N* expressed significantly lower than *SNORD119B* [F(2, 1)=9.08, $P=0.011$].

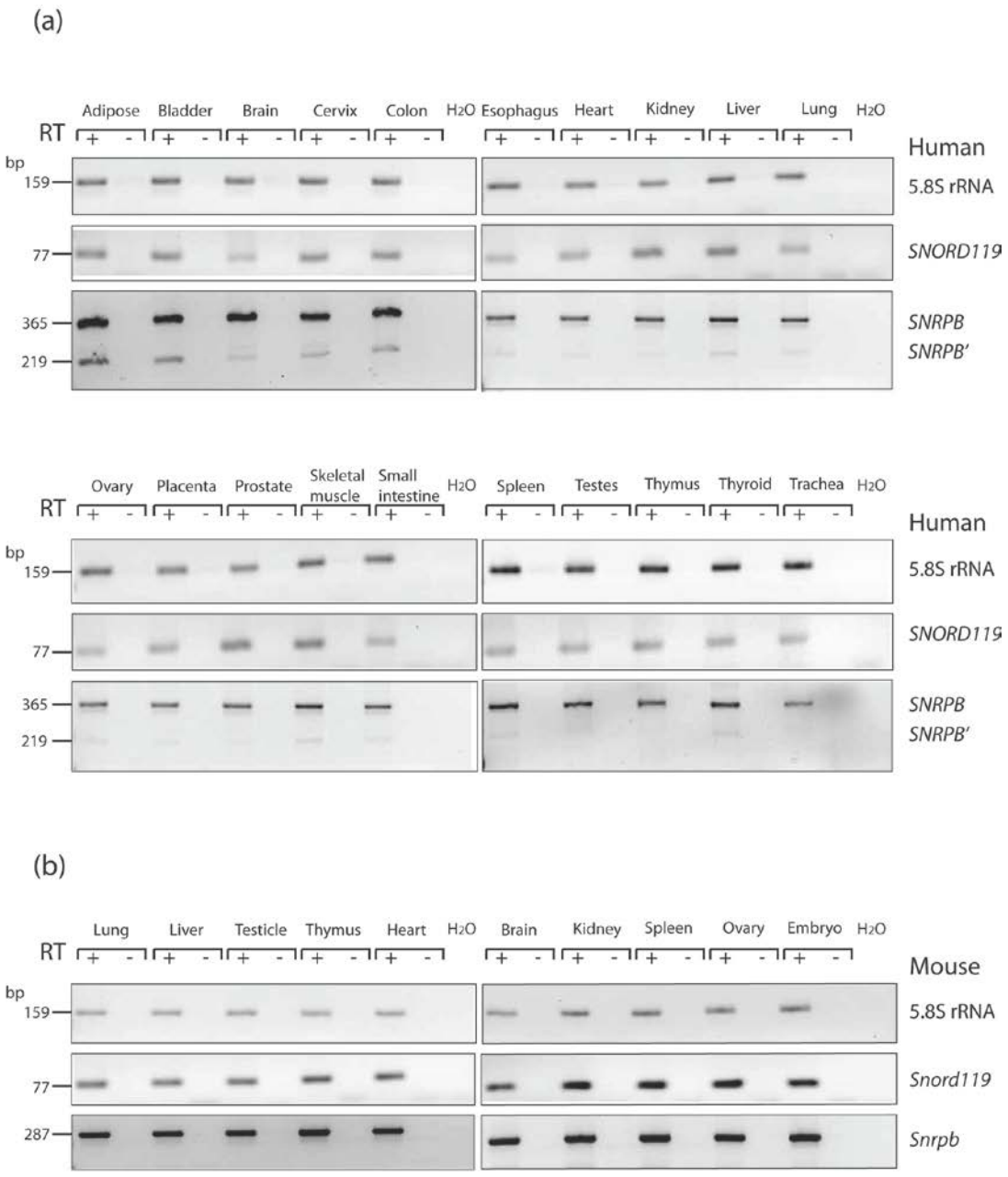


Figure 2.10. Expression studies of *SNORD119*, *SNRPN* and *SNPRB'* in human and mouse. (a) RT-PCR in 20 tissues from human. **(b)** RT-PCR in 10 tissues of mouse. Abbreviations: RT +: reverse transcription positive sample; RT -: reverse transcription negative control; bp: base pair.

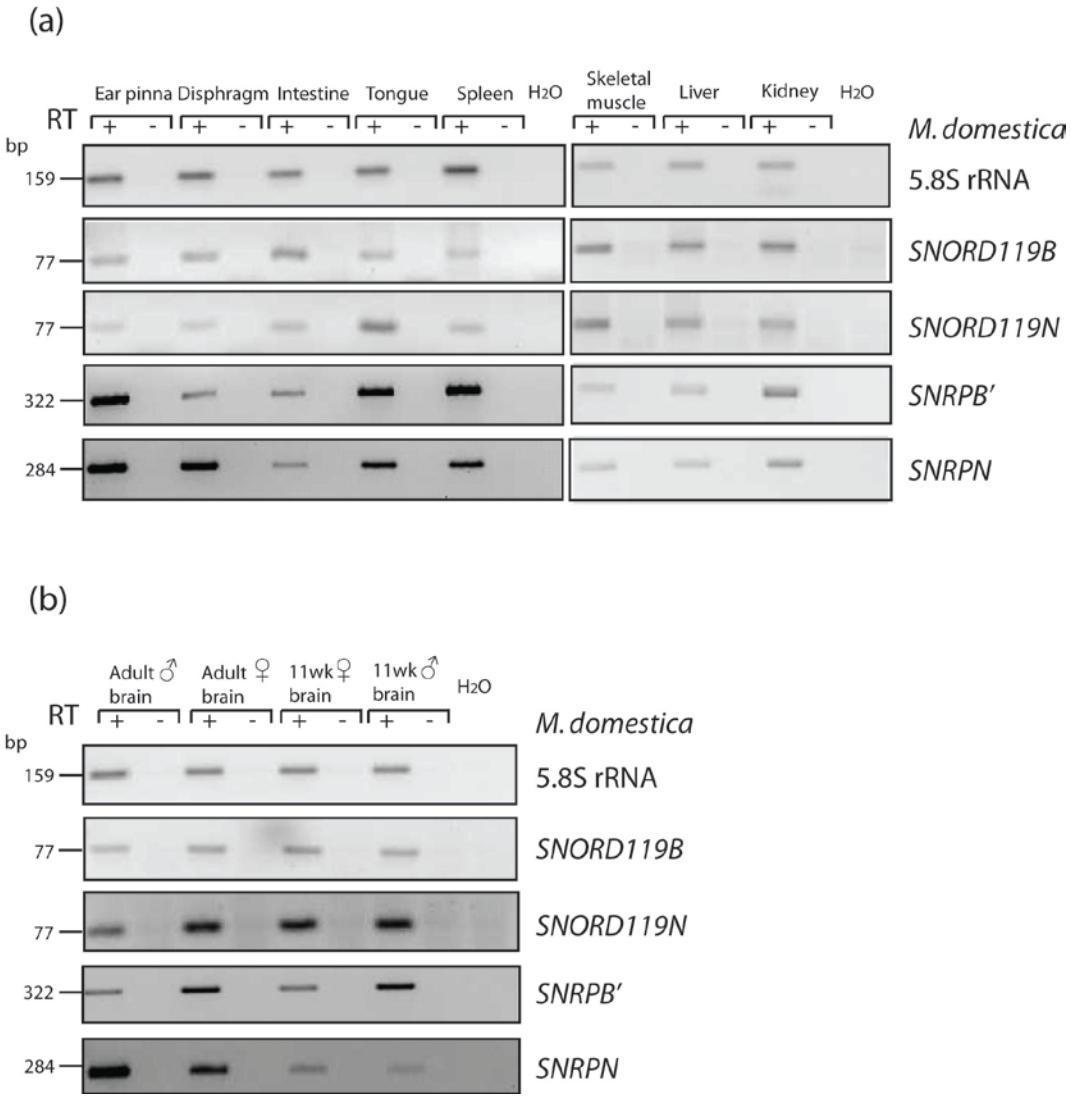


Figure 2.11. Expression studies of *SNORD119N*, *SNORD119B* and their host genes in *M. domestica*. (a) RT-PCR in 8 non-brain tissues of *M. domestica*. **(b)** RT-PCR in 4 brain tissues of *M. domestica*. Adult 1 ♂: animal #J7614, male; Adult 2 ♀: animal #J7344, female; 11 wk ♀: juvenile *M. domestica,* 11 weeks postpartum, female; 11 wk ♂: juvenile *M. domestica,* 11 weeks postpartum, male.

Figure 2.12. Verification of *SNORD119B* and *SNORD119N* snoRNA-specific RT-PCR assays by restriction enzyme fragment length variants (RFLV). (a) Alignment of the snoRNA sequences, with the position of RT-PCR primers in blue rectangles and the recognition site for the *Hae*III restriction enzyme in a red rectangle. (b) RFLV analysis on PCR products from RT-PCR using *SNORD119B*-specific and *SNORD119N*-specific primers. Abbreviations: Ta: annealing temperature in PCR; *Hae*III +: PCR product was digested by *Hae*III; *Hae*III -: PCR product was not digested, as a negative control.

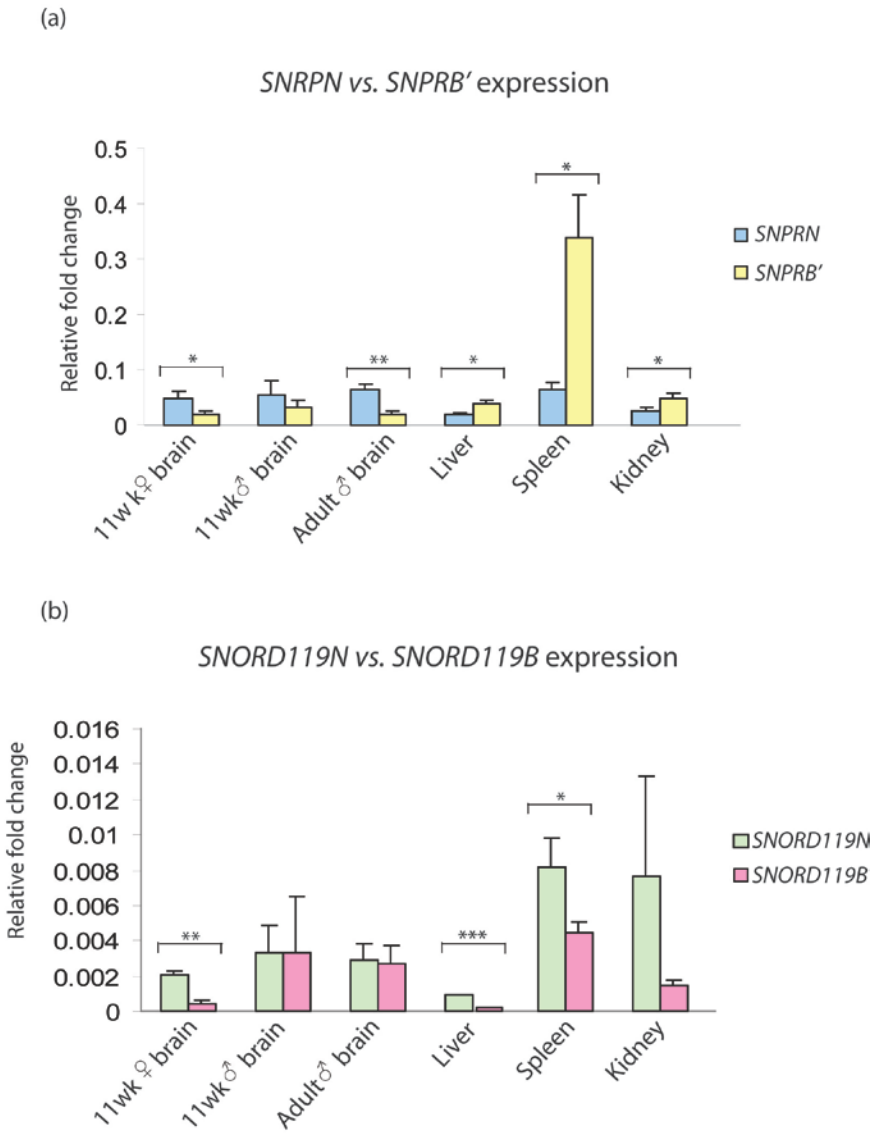


Figure 2.13. Expression studies of *SNORD119* snoRNAs by QRT-PCR in *M. domestica*. Abbreviations: Adult ♂: animal #J7614, male; 11 wk ♀: juvenile *M. domestica,* 11 weeks postpartum, female; 11 wk ♂: juvenile *M. domestica,* 11 weeks postpartum, male. Liver, spleen and kidney were from a female adult animal #G3232. Group comparisons of (a) *SNRPN* and *SNRPB'*; and (b) *SNORD119N* and *SNORD119B* were calculated by 2-sample t-tests using Microsoft Excel 2003. *P* values <0.05, <0.001, <0.0001 are represented by "*", "**" and "***", respectively.

2.3.3.2 Function of *SNORD119* orthologs and paralogs

As noted above, the C/D box snoRNAs have one or two 10-21 nucleotide long antisense elements 5' of box C/C', which are complementary to the mature rRNA and guide site-specific 2'-*O*-methylation. Structural conservation of the core of rRNA and modification is conserved between *Xenopus laevis* and human (Maden, 1990). In the same study, by oligonucleotide fingerprinting, the 4560 position of 28S rRNA was methylated in both human and *Xenopus laevis*; however, the same position was unmethylated in chicken 28S rRNA (Khan & Maden, 1976; Maden & Hughes, 1997, listed as position 4550 in their paper). When we aligned *SNORD119B* antisense box sequence to rRNA position 4560, we found the interaction was thermodynamically favorable (**Figure 2.14**). Similarly, *SNORD119N* antisense box was also thermodynamically favorable in aligning to the same position of rRNA in *M. domestica*. Therefore, we predicted the function of *SNORD119B* and *SNORD119N* is to target rRNA 4560 site for 2'-*O*-methylation (**Figure 2.14**). However, experimental studies are needed to test which snoRNA targets 28S rRNA *in vivo*. It is likely that *SNORD119B* is the common form and functions in most circumstances, while *SNORD119N* may function in the brain.

Similar targeting sequence was found in human 28S by *SNORD119* (**Figure 2.15**). Such interaction was thermodynamically favorable. Surprisingly, we found that SNORD115-3 was also thermodynamically favorable in potentially targeting the same position of 28S rRNA (**Figure 2.15**).

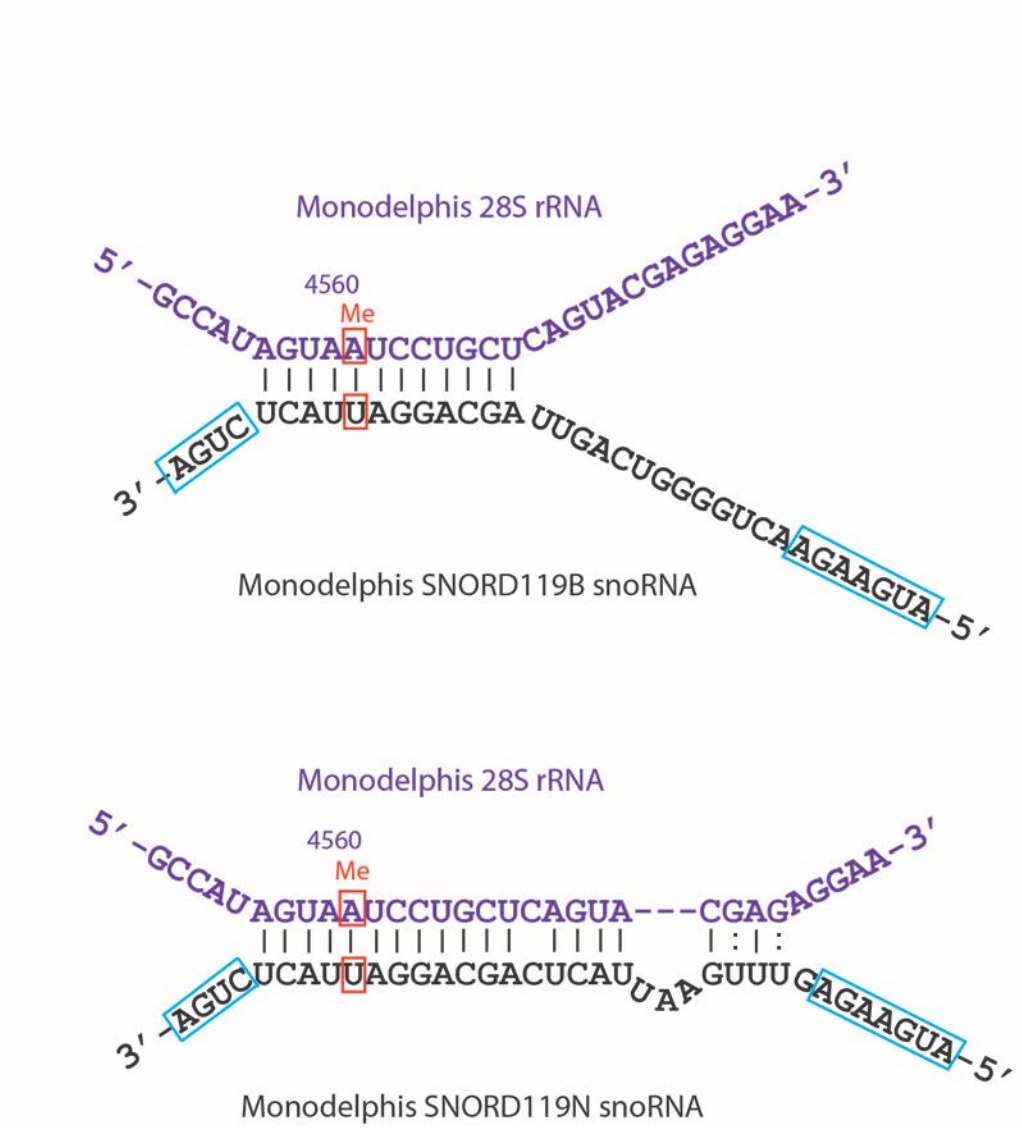


Figure 2.14. Predicted *M. domestica* 28S rRNA targeting by *SNORD119B* and *SNORD119N*. Box C' and box D are shown in blue rectangles. The predicted minimum free energy of *SNORD119B* binding to *M. domestica* 28S rRNA is ΔG°(37°C)=-23.6 kcal/mol, and for *SNORD119N* binding to *M. domestica* 28S rRNA is ΔG°(37°C)=-28.3 kcal/mol.

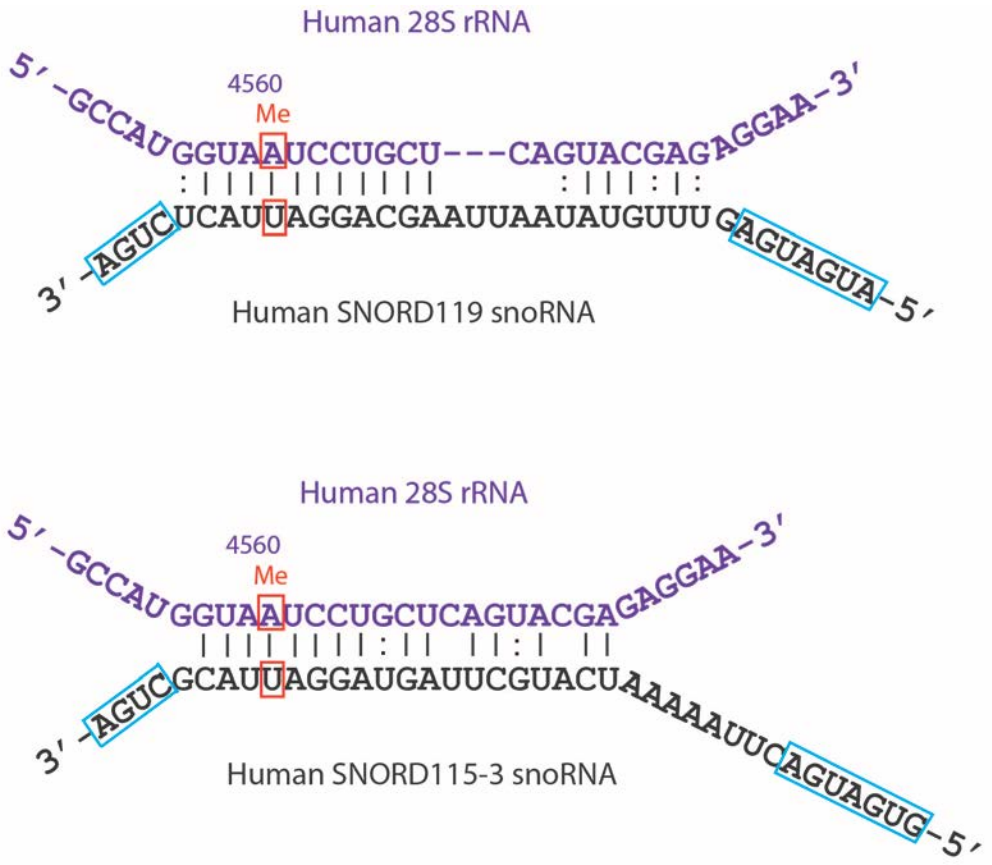


Figure 2.15. Comparison of human *SNORD119* and *SNORD115-3* for their potential targeting of 28S rRNA. Box C' and box D are shown in blue rectangles. The predicted minimum free energy of *SNORD119* binding to 28S rRNA is ΔG°(37°C)=-28 kcal/mol, and for *SNORD115-3* binding to 28S rRNA is ΔG°(37°C)=-26.9 kcal/mol.

2.3.3.3 Molecular phylogeny of *SNORD119* family descendants: the PWS imprinted snoRNAs

Unlike *M. domestica,* in human and mouse, *SNRPN* is located on a different chromosome to *SNRPB'* and forms a more complicated bicistronic locus, *SNURF-SNRPN,* in the PWS domain. The adjacent gene structure of paralogous *SNRPN* and *SNPRB'* only exists in metherians (both *M. domestica* and *D. virginiana*), but not in birds, reptiles, and platypus (data not shown), indicating the origin of the *SNRPN* occurred in a common ancestor of therians.

Downstream of the *SNURF-SNRPN* bicistronic locus in the PWS region, there are five classes of C/D-box snoRNAs in the intronic regions of extended host transcripts derived from alternative *SNURF-SNRPN* promoters (**Figure 2.16**). Among the five classes of PWS C/D-box snoRNAs, there are some copy number variances in different classes. *SNORD64* and *SNORD107* each has one single copy. *SNORD109* has two identical copies A and B. *SNORD115* has 47 copies. *SNORD116* has 29 copies (**Figure 2.16**, Cavaillé *et al*., 2000; Runte *et al*., 2001). In our analysis, we found that *SNORD64* was highly similar to *SNORD115*, *SNORD107* was highly similar to *SNORD109*, while *SNORD115* was also related to *SNORD109* or *SNORD107* (**Figure 2.17**). The three sub-classes of *SNORD116* are significantly more divergent to the other groups of PWS C/D-box snoRNAs. To better assess the sequence comparison and evolutionary studies between *SNORD119N* and PWS C/D-box snoRNAs, we BLASTed human *SNORD64*, *SNORD107* and *SNORD109* in the WGS database in NCBI and aligned maximum matched sequences from difference eutherian species to the human sequences, then achieved eutherian consensus sequences for *SNORD64*, *SNORD107*, and *SNORD109* (**Figure 2.18**). *SNORD119N* snoRNA showed identity to each class of PWS C/D-box snoRNAs (**Figure 2.19**). Combined, the information on the similarity between *SNRPN* versus the *SNURF-SNRPN* bicistronic locus and *SNORD119N* snoRNA versus the five classes of PWS C/D-box snoRNAs, suggested that the eutherian *SNURF-SNRPN*-

snoRNA locus evolved from a common ancestor as did the metatherian *SNRPN-SNORD119N* locus.

Based on these analyses, a gene evolutionary model (**Figure 2.20**) was established. In this model, we predicted that about 173-190 mya, the common ancestor of therians had only one copy of *SNRPB'*. The ancestral copy of *SNRPB'* was duplicated to form a paralogous gene, *SNRPN*, upstream of the ancestral *SNRPB'*. *SNRPN* then became fixed evolutionarily due to gain of functional requirement and selective pressure. In metatherians, the adjacent paralogs of *SNRPN* and *SNRPB'* remains in the same position. In a eutherian ancestor, however, *SNRPN* was translocated to another chromosome, and ontained *SNURF* by *de novo* gene formation or an unknown mechanism (Gray *et al*., 1999a; Chai *et al*., 2001). During the evolution from eutherian ancestor to modern eutherians, other chromosomal rearrangement occurred. In our model, combined with other evolutionary studies, *MKRN3* was acquired by retrotransposition from an ancestral *MKRN1* gene (Gray *et al*., 2000, 2001), and the intronless *NDN* and *MAGEL2* were acquired by retrotransposition from an ancestral *MAGE* gene on the X chromosome and subsequent duplication (Chai *et al*., 2001; Rapkins *et al*., 2006). Subsequent to multiple duplication and diversification events, the *SNORD119* gene gave rise to the five classes of PWS C/D-box snoRNAs, although these have unknown functions. Additionally, the *SNORD115* and *SNORD116* C/D-box snoRNA families have duplicated many times to form different sub-classes (**Figure 2.21 & Figure 2.2).**

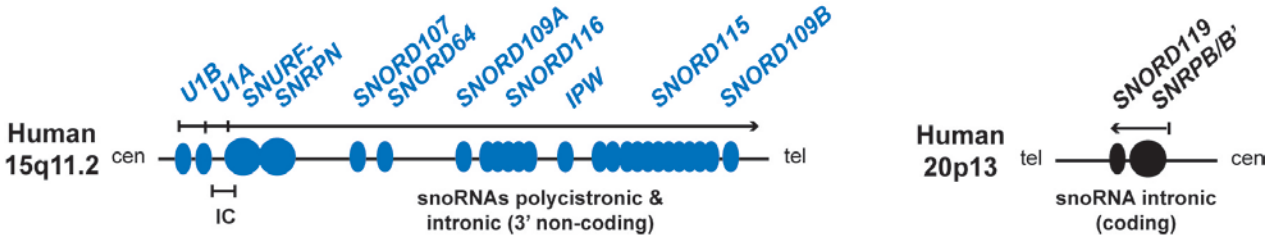


Figure 2.16. Map of PWS C/D-box snoRNAs. Circles represent protein coding genes, and ovals represent RNA genes. Abbreviations: IC: imprinting center; cen: centromere; tel: telomere.

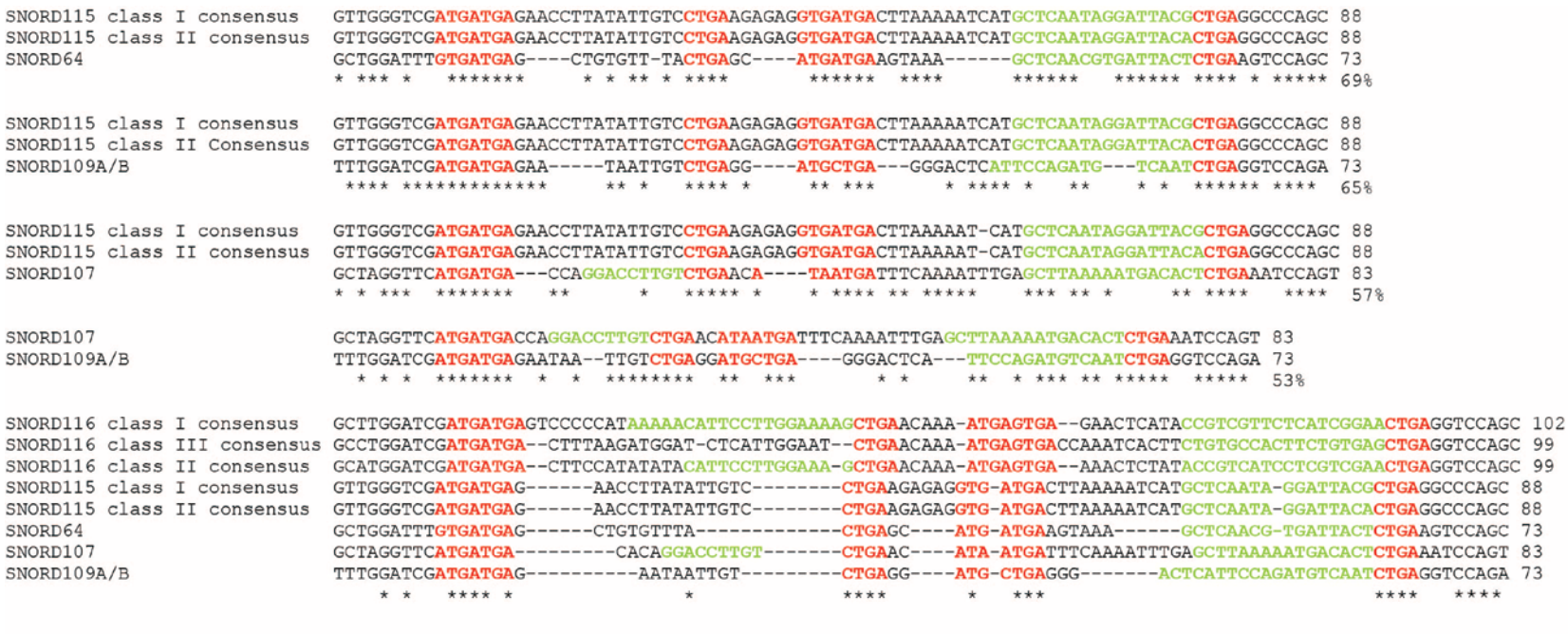


Figure 2.17. Comparison of five classes of PWS C/D box snoRNAs. Box C/C' and box D/D' are highlighted in bold red: from 5' to 3', these are box C, box D', box C' and box D, respectively. The antisense box is highlighted in bold green and is 5' of box D.

(a)

SNORD64 eutherian orthologs alignment and consensus sequence

```
Chimpanzee                    GCTGGATTTGTGATGAGCTGTGTTTACTGAGCATGATGAAGTAAA-GCTCAACGTGATTACTCTGAAGTCCAGC 73
Sumatran orangutan            GCTGGATTTGTGATGAGCTGTGTTTACCGAGCATGATGAAGTAAA-GCTCAACGTGATTACTCTGAAGTCCAGC 73
Human                         GCTGGATTTGTGATGAGCTGTGTTTACTGAGCATGATGAAGTAAA-GCTCAACGTGATTACTCTGAAGTCCAGC 73
Rhesus monkey                 GCTGGATTTGTGATGAGCTGTGTTTACTGAGCATGATGAAGTAAA-GCTCAATGTGATTACTCTGAAGTCCAGC 73
Small-eared galago            GTTGGATTTATGATGAGCTGTGTTTAGTGAGCATGATGAAGTAAA-GCTCAATGTGATTACTCTGAAGTCCAGC 73
Gray mouse lemur              GCTGGATTTATGGTGAGCTGTGTTTACTGAGCATAATGAAGTGAA-GCTCAAAATGATTAATCTGAAGTCCAGC 73
Rat                           GCTGGATTAATGATGAGCTGTGTTTACTGAGCATAATGAAATGAC-GCTCAAAATAATTACTCTGAAGTC-AGC 72
Rabbit                        GCTGGATTTATGATGAGCTGTATTTACTGAGCATAATGAAGGGAG-GCTCCATGTGATTACTCTGAAGTCCAGC 73
Thirteen-lined ground squirrel GCTGGATTTATGATGAGCTATGTTTACTGAGCATGATGAAGTGAA-GCTCAATATGATTACTCTGAGGTCCAGC 73
Western European hedgehog     GCTGGATTTATGATGAGCTATGTTTACTGAGCATGATGAAATGAA-GCTTGATATGAGAACTCTGAGGTCCAGT 73
Mouse                         GCTGGACTTATGATGAGCTATGTTTACTGAGCATGATGAAATGAT-GCTCAAAATGATTACTCTGAAGTC-AGC 72
Nine-banded armadillo         GCTGGATTTATGATGAGATATGTTTACTGAGCCTGATGAAAAGAAAGCTCACTGTGATTACTCTGAAGTCCAGC 68
Cat                           ACTGGATTTATGATGAGTTATGTTTACTGAGTATGATGAAGTGAA-ACTCAGTGTGAGTACTCTGAAGTCCAGC 67
                                ******  ** **** * * **** ****  * *****   *   **     *** ************ **

Consensus                     GCTGGATTTRTGRTGAGCTRTGTTTACTGAGCATRATGAARTRAA-GCTCAAHRTGATTACTCTGAAGTCCAGC 73
```

(b)

SNORD107 eutherian orthologs alignment and consensus sequence

```
Rhesus monkey                 TTGGCTAGGTTCATGATGACACAGGACCTTGTCTGAAC-ATAATGATTTCAAAATTTGAGCTTAAAAATGACACTCTGAAATCCAGTCAG 89
Human                         TTGGCTAGGTTCATGATGACACAGGACCTTGTCTGAAC-ATAATGATTTCAAAATTTGAGCTTAAAAATGACACTCTGAAATCCAGTCAG 89
Chimpanzee                    TTGGCTAGGTTCATGATGACACAGGACCTTGTCTGATC-ATAATGATTTCAAAATTTGAGCTTAAAAATGACACTCTGAAATCCAGTCAG 89
Sumatran orangutan            TTGGCTAGGTTCATGATGACATAGGACCTTGTCTGAAC-ATAATGATTTCAAAATTTGAGCTTAAAAATGACGCTCTGAAATCCAGTCAG 89
Dog                           TTGGCTAGGTTCATGATGACATAGGACCTTGTCTGAAC-ATAATGA-TTCAAAATTTGAGCTTAAA--TGACACTCTGAAATCCAGTCA- 85
Small-eared galago            TTGGCTAGGTTCATGATGACATAGGACCTTGTCTGAAC-ATAATGATTTCAAAACTTGAGCTTAAAA-TGACA--CTGAAAGCCAGTCA- 85
Cat                           TTGGCTAGGTTCATGATGACACAGGACCTTGTCTGAAC-ATAATGATTTTAAAATTTGAGCTTAAA--TGACACTCTGAAATCTAGTCA- 86
Thirteen-lined ground squirrel TTGGCTAGGTTCATGATGACACAGGACCTTGTCTGAAT-ATGATGATTTCAAAATTTGAGCTTAAA-TTGACACTCTGAAATCTAGTCAG 88
American pika                 TTGACTAGGTTCATGATGACACAGGACCTTGTCTGAAAATTAATGATTTCAAAATTTGAGCTTAAAA-CGACATTCTGAAATCTGGTCA- 88
Little brown bat              -TGGCTAGGTTCATGATGACACAGGACCTTGTCTGAAC-TTCATGATCTCAAAACATGAGCTTGGATATGTCA--CTGAAATCTAGTCA- 85
Horse                         TTGGCTAGGTTCATGATGATACAGGACCTTGTCTGAAC-AAAGTGATTTCAAAATTTGAGCTTAAAA-TGACATTCTGAAATCCAGTCA- 87
Northern tree shrew           TTGGCTAGGTTCATGATGACACTGGACCTTGTCTGAAA-AAAGTGATTTCAAAATTTGAGCTTAAAA-TAATACTCTGAAATCCAGTCA- 87
European shrew                TTGGCTAGATTAATGATGATACAGGACCTTGTCTGAAT-ATACTGATTTCAAAATTTGAGCTTAAAA-TAATATTCTGAAATCCAGTCA- 87
Nine-banded armadillo         TTGGCTAGATTCATGATGACACAGGACCTTGTTTGAAA-ATAATGATATTAAACTTTGAGCTTAAA--CCACACTCTGAAATCCAGTC-- 85
Western European hedgehog     TTGGCTAGATTCATGATGACATTGGACCTTGTATGACT-AAAATGACATCAAAATTTGAGCTTAGAA-TTATACTCTGAAATCCAGTCA- 87
Rabbit                        TTGGCTAGGTTCATGATGACATAGGACCTTGTCTGAATAATAATGATTTCAAATTTTGAGCTAAAG--TGACTTTCTGAAATCCAGTCAG 88
Mouse                         TTGACTGGATTTATGATGATATAGGACCTTGTCTGAAT-ATAATGATTTCAAAAATTGAGCTTACA---GTGA--CTGAAATCTAGTCAA 84
Rat                           TTGACTAGATTTATGATGACATGGGACCTTGTCTGAAC-ACAATGATTTCAAAAATTGAGCTTAGA---GTGACTCTGAAATCTAGTCA- 85
                               ** ** * ** ******* *  ********* ***       ***  * ***   ******             ****** *  ***

Consensus                     TTGRCTAGRTTYATGATGAYAYAGGACCTTGTCTGAAn-AWAATGATTTYAAAAnTGAGCTTAAAAATGACACTCTGAAATCYAGTCAG 89
```

(c)

SNORD109 eutherian orthologs alignment and consensus sequence

```
Human                         TTTGGATCGATGATGAGAATAATTGTCTGAGGATGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 73
Rhesus monkey                 TTTGGATCGATGATGAGAATAATTGTCTGAGGATGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 73
Chimpanzee                    TTTGGATCGATGATGAGAATAATTGTCTGAGGATGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 73
Sumatran orangutan            TTTGGATCGATGATGAGAATAATTGTCTGAGGATGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 73
American pika                 TCTGGATCGATGATGAGAAAAAATGTCTGAGGATGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 71
Cattle                        TCTGGATCGATGATGAGAAAAACTGTCTGAGGATGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 71
Rabbit                        TCTGGATCGATGATGAGAACAAGTGTCTGAGGATGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 71
Small-eared galago            TCTGGATCGATGATGAGCAAAATTGTCTGAGGACGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 71
Gray mouse lemur              TCTGGATCGATGATGAGCAAAATTATCTGAGGACGGTGAACGACACATTGCAGATGTCAATCTGAGTTCCAGA 71
European shrew                TCTGGATCAATGATGAGAAAAATTGTCTGAGGACGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 71
Horse                         TCTGGATCAATGATGAGCATAATTGTCTGAGGATGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 71
Nine-banded armadillo         TCTTGATCAATGATGAGCATAATTGTCTGAGGATGCTGAGGGACTCATTCTAGATGTCAATCTGAGGTCCAGA 69
Domestic guinea pig           TCTGGATCAATGATGAGCACAATTGTCTGAGGATGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 71
Dog                           TCTGGATCAATGATGAGCACAATTGTCTGAGGATGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 71
African savanna elephant      TCTGGGTCAATGATGAACATAAATGTCCGAGGATGCTGAGTGACTCATTGCAGATGTCAATCTGAGATCGAGA 71
                              * * * ** *******    ** * ** ***** * ***  *** **** ************** ** ***

Consensus                     TYTGGGTCRATGATGAGMAnAATTGTCTGAGGAYGCTGAGGGACTCATTCCAGATGTCAATCTGAGGTCCAGA 73
```

Figure 2.18. Multisequence alignments of *SNORD64*, *SNORD107*, and *SNORD109*. Box C/C' and box D/D' are highlighted in bold red: from 5' to 3', these are box C, box D', box C' and box D, respectively. The antisense box is highlighted in bold green and is 5' of box D.

Figure 2.19. Comparison of *SNORD119N* to five classes of PWS C/D-box snoRNAs. Box C/C' and box D/D' are highlighted in bold red: from 5' to 3', these are box C, box D', box C' and box D, respectively. The antisense box is highlighted in bold green and is 5' of box D.

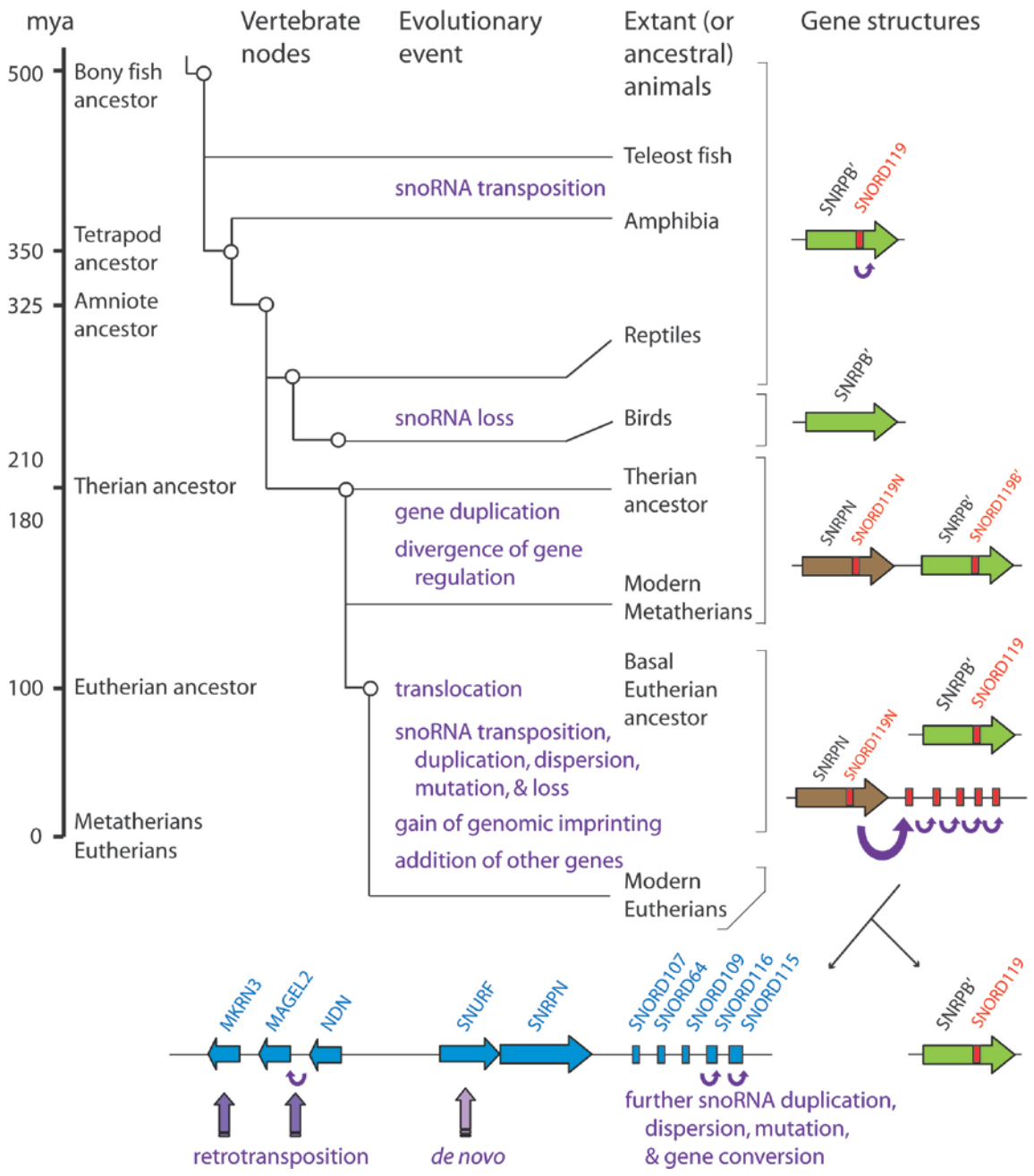


Figure 2.20. Evolutionary model for the origin of the PWS imprinted domain.

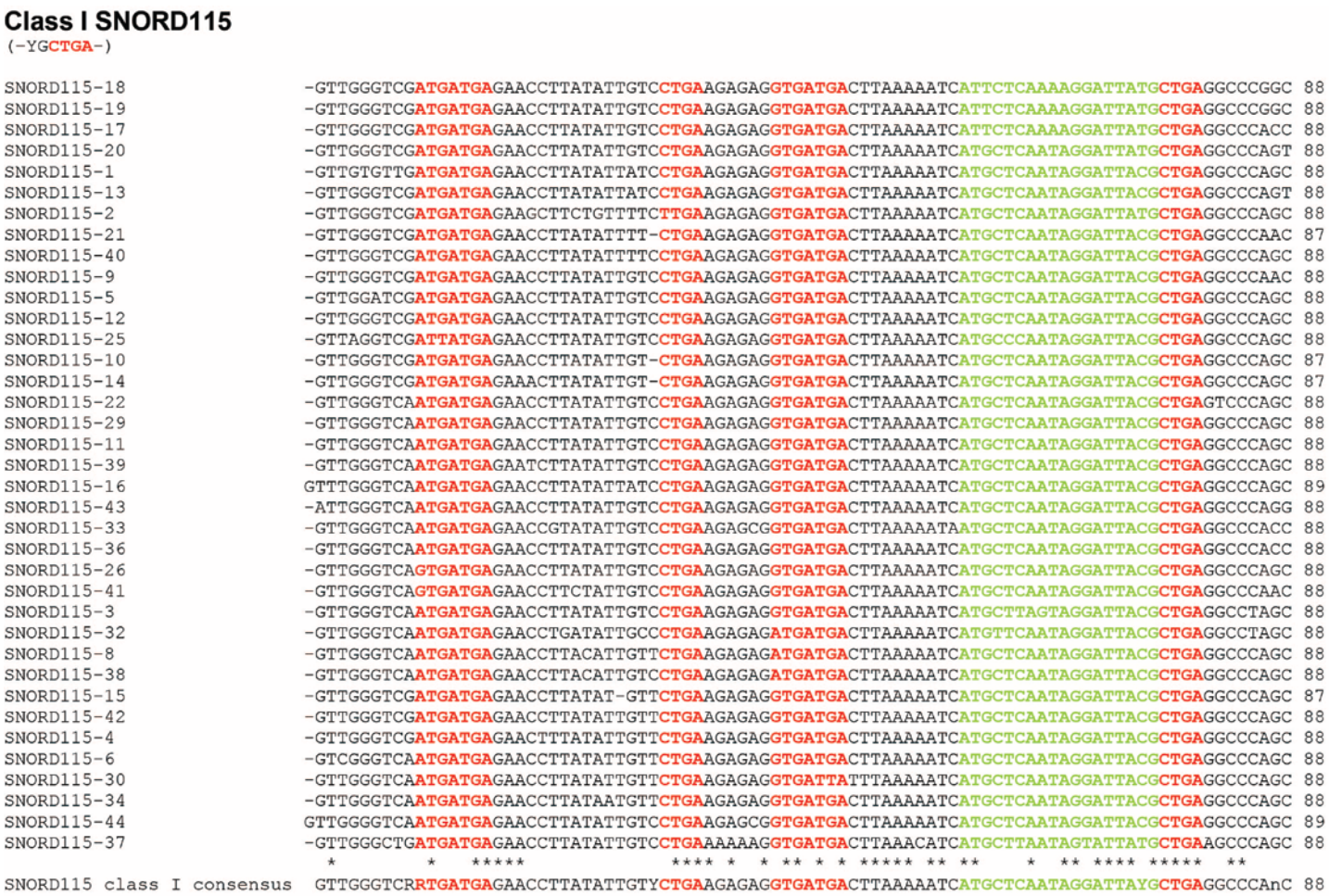


Figure 2.21. (continued below)

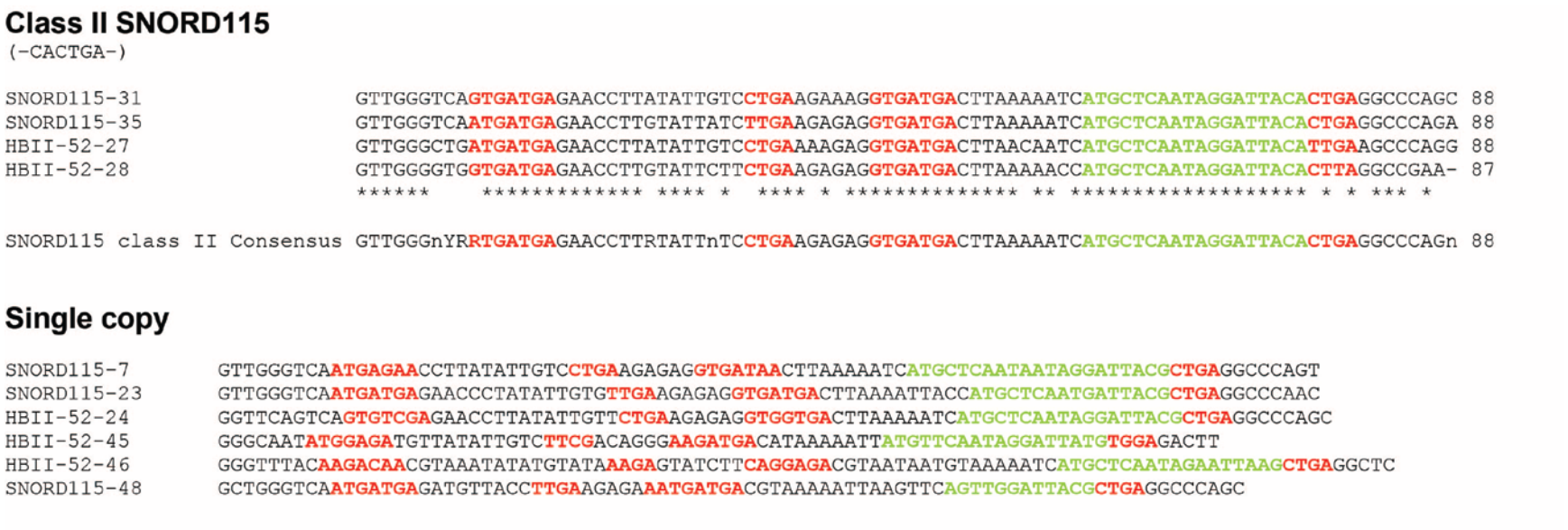


Figure 2.21. Human *SNORD115* family sequence alignment and consensus sequence of each class. Box C/C' and box D/D' are highlighted in bold red: from 5' to 3', these are box C, box D', box C' and box D, respectively. The antisense box is highlighted in bold green and is 5' of box D.

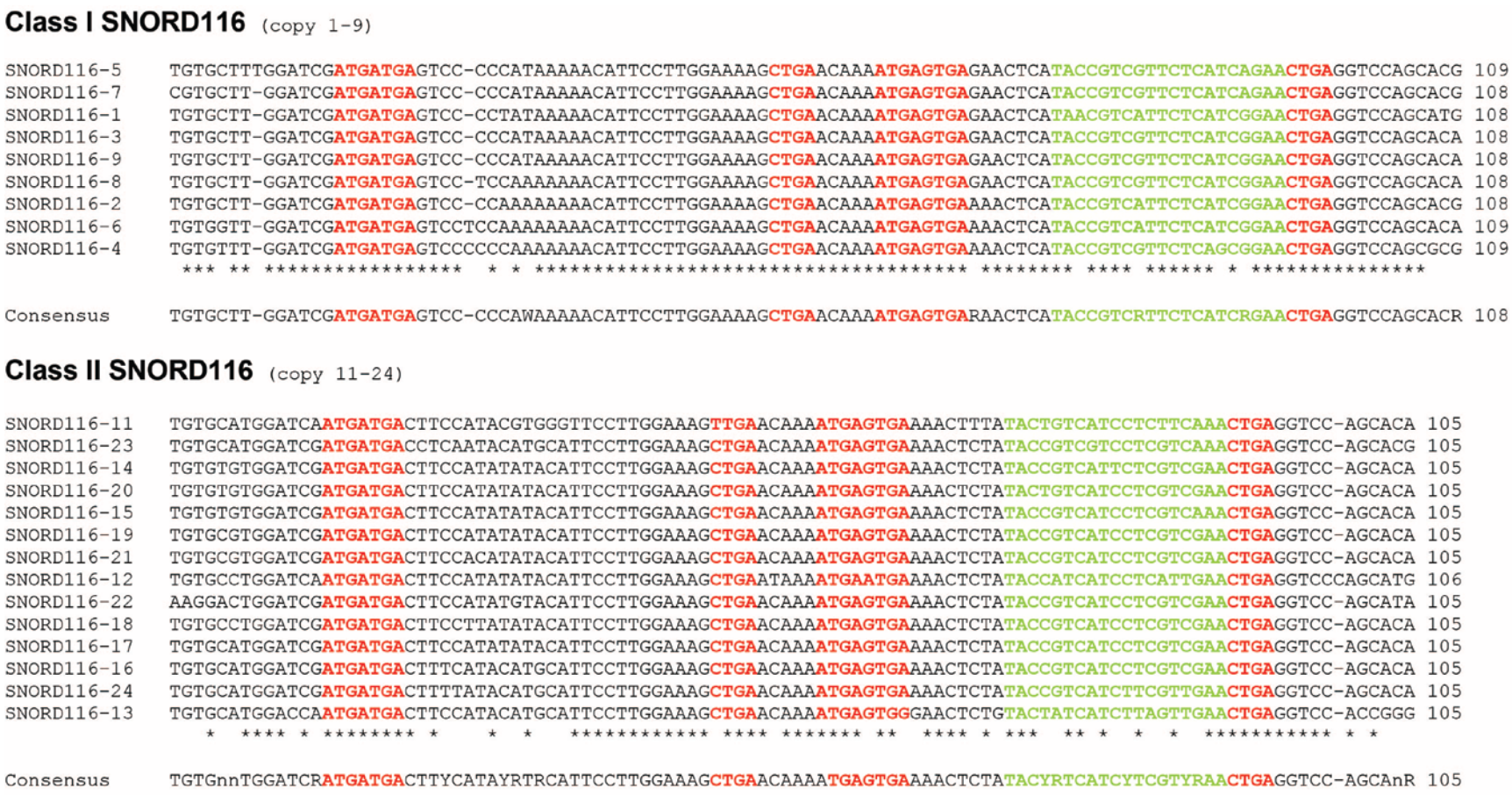


Figure 2.22. (continued below)

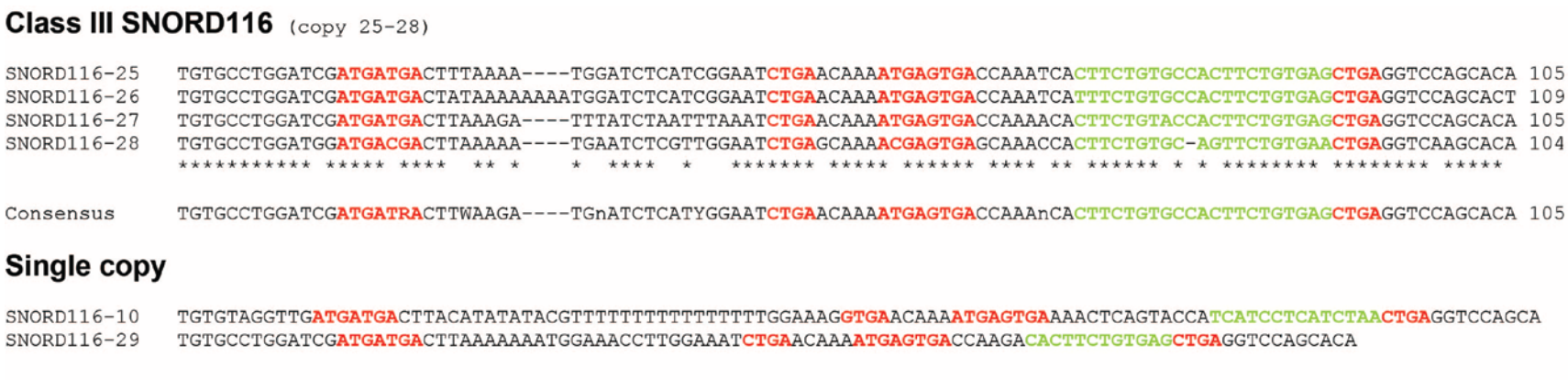


Figure 2.22. Human *SNORD116* family sequence alignment and consensus sequence of each class. Box C/C' and box D/D' are highlighted in bold red: from 5' to 3', these are box C, box D', box C' and box D, respectively. The antisense box is highlighted in bold green and is 5' of box D.

2.4 DISCUSSION

2.4.1 NRF1 regulation of the PWS imprinted domain in somatic cell hybrids

By ChIP assays, we have confirmed the active and repressive chromosomal domains in GM11715, A15 and A15-1 cells. These data showed weak binding of NRF1 to the "NRF1 cluster" in these three somatic cell hybrids. Since our goal was to use a somatic cell hybrid with strong binding of NRF1 to the "NRF1 cluster" in the single copy of human chromosome 15 to study NRF1 regulation on the PWS region, these results indicated none of tested cell lines would be ideal for the "gene targeting" study of "NRF1 cluster". However, my data provided a valuable resource of chromatin information on these three somatic hybrid cell lines for other work in our laboratory.

2.4.2 Evolutionary insights and relationships of the PWS *SNRPN* imprinted locus and snoRNAs

In this study, we reported novel snoRNAs in the *SNRPN* and *SNRPB'* genes of *M. domestica*, *SNORD119N* and *SNORD119B*, respectively, and suggested that by gene duplication of *SNRPB' to SNRPN, SNORD119N* was retained and duplicated from *SNORD119B* in intron 5 of *SNRPB'*. In eukaryotes, functional related genes are in clusters and can be derived from tandem duplication of an ancestral gene (Spitz & Duboule, 2006). Synteny between two species indicates the gene regions have been evolved from a common ancestral gene region (Tanaka-Fujita *et al*.,

2007). For snoRNAs, duplication of a host gene can generate snoRNA paralogs (Weber, 2006, Zemann *et al*., 2006). Combining the sequence similarity of *M. domestica SNORD119N* versus each of the five classes of PWS C/D-box snoRNA genes, the sequence similarity among the five classes of PWS C/D-box snoRNAs, and the syntenic regions of *SNRPN* and *SNRPB'* between human and *M. domestica*, we concluded that *SNORD119N* in *M. domestica* is derived from the same common ancestral snoRNA as for the five classes of PWS C/D-box snoRNA genes by multiple rounds of DNA duplications (**Figure 2.19**). Similar DNA duplication events have been found with multiple copies of U1 promoters in the PWS region (Färber *et al*., 1999). In our analysis, *SNORD119N* is more related to *SNORD64* and *SNORD115* (**Figure 2.18**). This suggests either a more recent evolutionary origin of these snoRNAs or possibly retention of similar functions. Nevertheless, the large number of duplication and diversification events in forming the eutherian PWS-region snoRNAs suggests dynamic evolutionary processes operate on this domain and have selected for evolutionarily. It will be necessary to determine the function of each of the PWS-region snoRNAs in order to understand what evolutionary forces might have driven these processes.

Because *SNURF-SNRPN* is associated with the imprinting control region in the PWS region, our phylogenetic study on the origin of *SNRPN* will have important implications in understanding gene functions and the imprinting mechanism in the PWS region. The comparison of eutherian genomes and the metatherian genome (*M. domestica*) is highly effective in recognizing genes and regulatory sequence evolution. Compared to the *M. domestica* genome, higher mammals have evolved by modification of epigenetic mechanisms instead of primarily inventing new genes (Mikkelsen *et al*., 2007; Renfree, 2007). Genomic imprinting has been found in both eutherians and marsupials, but seems to have a different imprinting mechanism in

marsupials. To better understand why and how imprinting was established in phylogeny, studies on the imprinting mechanism in marsupials become critical. It has been described that the origin of *SNRPN* was derived from gene duplication of an ancestral *SNRPB'* gene, and the expression of brain-specific-imprinted SmN could be compensated by a feed-back loop of SmB/B' expression in order to maintain stoichimetric levels of spliceosomal components (McAllister *et al*., 1989; Gray *et al*., 1999b; Matera *et al*., 2007). The indispensability of SmB'/B/N was consistent to our finding that *SNRPB'* existed in all species (data not shown), while *SNORD119* existed in most species we analyzed, including all mammals, one amphibian, reptiles, and two kinds of fish. The constrained sequence of *SNRPB'* and *SNORD119* implied the functional importance of both genes.

In this study, we also predicted the function of *SNORD119* is to target 28S rRNA at position 4560 for 2'-*O*-methylation based on sequence alignment and thermodynamic calculation (**Figure 2.14; Figure 2.15**). Methylation modification of rRNA on position 4560 has been experimentally confirmed in human and *X. laevis*, but was not found in chicken (Khan & Maden, 1976; Maden & Hughes, 1997, listed as position 4550). Consistent with this, *SNORD119* was found in *SNRPB'* of human and *X. laevis*, but not in chicken.

Considering evolutionary explanation for the snoRNA copies in PWS region, it can be traced by the sequence similarity between *M. domestica SNORD119N* and each of the five classes PWS C/D-box snoRNAs. First, PWS C/D-box snoRNAs were derived from *SNORD119N* by a series of DNA duplications after which they gained new function in brain neonatal and/or postnatal development. This is supported by recent studies that snoRNAs can also target pre-mRNA modification for gene silencing and regulation (Zhao & Yu, 2008) and *SNORD115* can target serotonin 2C receptor to regulate gene expression (Kishore & Stamm, 2006). There is a

growing number of snoRNA found to have tissue-specific expression pattern, most probably reflecting that of the host gene (Cavaillé *et al*., 2000; Runte *et al*., 2001). A PWS patient was found to have distal *SNURF-SNRPN* translocation breakpoint with no gene expression of *SNORD116*, *IPW* and *PAR1* (Wirth *et al*., 2001). A mouse model with a paternal deletion from *Snrpn* to *Ube3a* showed a phenotype of hypotonia, growth retardation, and partial lethality (Tsai *et al*., 1999) that is similar to the neonatal presentation of PWS and other PWS mouse models. These data suggest that the loss of PWS C/D-box snoRNAs contributes to the disease of PWS (Cavaillé *et al*., 2000; Runte *et al*., 2001). Another recent study on a paternally-inherited *Snord116* deletion mouse model showed that snoRNA deletion in the PWS region was not lethal, but led to growth deficiency and a mild hyperphagia (Ding *et al*., 2008) albeit the animals weighed less than control littermates. All these studies suggest there is some specific function of PWS C/D-box snoRNAs in neonatal to postnatal development. Alternatively, the PWS C/D-box snoRNAs that have high sequence similarity to *SNORD119N*, such as *SNORD64* and *SNORD115*, may target the same position of 28S rRNA in the brain. For example, in our analysis, we predicted that PWS C/D-box snoRNAs *SNORD115-3* could alternatively target the same position of 28S rRNA (**Figure 2.15**) during postnatal development for brain-specific function (Cavaillé *et al*., 2000; Runte *et al*., 2001; de los Santos *et al*., 2000).

Additional analysis, in our studies, implied some of the copies in *SNORD115* and *SNORD116* may not be functional because of mutations in the box-C/C' and box-D/D' regions (**Figure 2.21 & 2.22**), the gene structure of which were similar to the pseudogene *HBII-437* (Runte *et al*., 2001).

To summarize, our study identified the origin of the five classes PWS C/D-box snoRNAs by gene duplication and diversification from an ancestral host gene with its intronic snoRNA.

Clearly, snoRNA evolution is labile, with duplication, transposition, and loss being detected in different phylogenetic lineages. It remains unclear why multiple copies of PWS C/D-box snoRNAs exist. It may be due to dosage requirement in specific period of development and/or functional compensation to some loss-of-function copies, such as the ones in the classes of *SNORD115* and *SNORD116.* In the long term, functional studies on the five classes of PWS C/D-box snoRNAs will allow these questions to be addressed, which will also shed light on the role of the box C/D-snoRNAs in disease mechanisms of PWS.

3.0 REGULATORY ROLES AND MECHANISMS OF NRF1 IN CIRCADIAN RHYTHMS

3.1 INTRODUCTION

Circadian rhythms, an "internal body clock", govern the metabolic, hormonal, physiological, and behavioral pace of an organism in a 24-hour period, and exist from lower organisms such as bacteria to higher mammals such as human (Roenneberg & Merrow, 2005). The central mammalian circadian pacemaker located in the suprachiasmatic nucleus (SCN) of the hypothalamus is entrained by light acting on retinal neurons that project to the SCN, and uses neuronal and hormonal mechanisms to coordinate clocks in organs throughout the body (Maywood *et al*., 2007; Takahashi *et al*., 2008). In addition to the SCN, peripheral organs, such as liver and heart, have circadian rhythms that receive output signals from the SCN (Rutter *et al*., 2002; Hastings *et al*., 2007) and additionally can be entrained by food or energy balance (Mendoza, 2007).

At the cellular level, circadian rhythms are comprised of positive and negative feedback loops. In this feedback system, the CLOCK and BMAL1 polypeptides form a heterodimeric complex which binds to an E-box (5'-CANNTG-3') transcriptional element and functions as a transcriptional activator in the circadian pacemaker to regulate clock-controlled genes (CCGs, Steeves *et al*., 1999; Gekakis *et al*., 1998; Kondratov *et al*., 2003). In contrast, PER and CRY

polypeptides also form heterodimers and mediate the negative limb of the circadian transcriptional feedback loop (Sun *et al*., 1997; Darlington *et al*., 1998; Hastings *et al*., 2007; Brown *et al*., 2005). The increase of PER/CRY heterodimers inhibits CLOCK/BMAL1 function. Subsequently, PER and CRY are degraded and cycling of this system generates the circadian oscillations (Reppert & Weaver, 2002). Additionally, *NR1D1* forms an ancillary loop to coordinate the pacemaker (Takahashi *et al*., 2008) regulating *BMAL1* via *REV-ERBα* response elements (ROREs, Preitner *et al*., 2002; Ueda *et al*., 2002) and also to repress itself by autoregulation (Adelmant *et al*., 1996). The fine-tuned regulation of both loops leads CCGs regulated by CLOCK/BMAL1 to act under circadian control. Although many key regulatory genes in the circadian system have been found, there may still be some missing factors that coordinate the system.

It has been suggested that *DBP*, a key clock output gene, plays a role in the feedback system for circadian rhythms (Takahashi *et al*., 2008). *DBP* contributes to activation of *PER1* by binding to the *PER1* promoter (Yamaguchi *et al*., 2000). Expression of *DBP* is driven by oscillations of chromatin structure that regulate the binding of CLOCK/BMAL1 to E-box motifs within enhancer regions located in the first and especially the second intron (Ripperger *et al*., 2000; Ripperger & Schibler 2006). Indeed, deletion of two E-box motifs in intron 2 abolished the oscillation of *DBP* expression and chromatin structure in Rat-1 cells and in a mouse model (Ripperger *et al*., 2000; Ripperger & Schibler 2006). In this work, we identify a conserved NRF1 motif within the intron 2 enhancer element of *DBP*. Furthermore, we identify conserved NRF1 binding sites in the 5' regulatory elements of 25 of 45 circadian regulatory genes, including *CLOCK*, *CRY1*, *PER1*, and *NR1D1*. These are functional binding sites as shown by ChIP, expression studies following siRNA knockdown of *NRF1* mRNA levels, and transient

transfection assays of wildtype and mutant promoter and promoter-enhancer luciferase reporter constructs. Further, co-immunoprecipitation shows that NRF1 and the phosphorylated, active form of CLOCK interact in a molecular complex. In serum-induced NIH3T3 cells with circadian oscillations of *Dbp* and *Nr1d1* mRNA, *Nrf1* mRNA and protein levels show ultradian oscillations. Based on these results, we propose that NRF1 has a major role in activating expression of circadian regulatory genes from both the positive and negative feedback regulatory loops, and consequently is positioned to integrate many aspects of circadian biology.

3.2 MATERIALS AND METHODS

3.2.1 Bioinformatics and phylogenetic analyses

Genomic sequences for *CLOCK*, *PER1*, *CRY1*, *DBP*, *NR1D1*, and 45 other circadian regulatory genes were obtained by NCBI BLAST (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) with the human genes from Ensembl database. Sequences spanning each gene were exon-masked and/or Repeatmasked [using the "human" in "DNA source" option] and used for BLAST. The consensus NRF1 recognition sequence "YGCGCANGCGCR" were used for the search and allowed 1 nucleotide change in either "GCGC" arm. Putative NRF1 binding sites were searched 5 kb upstream of TSS and at least 1 kb into intron 1 of 45 reported circadian regulatory genes. For *CLOCK*, *PER1*, *CRY1*, *DBP*, and *NR1D1*, all sequence was transformed into FASTA format on ReadSeq (http://searchlauncher.bcm.tmc.edu/seq-util/readseq.html) and the alignments were performed by Clustal W (http://www.ebi.ac.uk/Tools/clustalw2/index.html). Alignments were adjusted as needed to maximize parsimony.

3.2.2 Gene expression studies

RNA was extracted from SK-N-SH human neuroblastoma cells and NIH3T3 mouse embryonic fibroblast cells, and reversed transcribed into cDNA using SuperScript® III reverse transcriptase according to the manufacturer's instruction (Invitrogen). Regular PCR was performed to test gene expression of 25 circadian regulatory genes using the primers in **Table 3.1**.

Table 3.1. RT-PCR primers for circadian gene expression studies (human SK-N-SH cells).

Target gene (primer)[a]	Primer number	Exon	Primer sequences
CLOCK (F)	RN3519	17	5′-GATAATCGTATAAACACAGTCAGTC-3′
CLOCK (R)	RN3520	17-18	5′-GGTGTTGAGGAAGGGTCTGAG-3′
PER1 (F)	RN3521	16	5′-GTCTCTGTGGGGACCAAGAAAG-3′
PER1 (R)	RN3522	17	5′-ACTGGCTCCTTCCGTGGGGTG-3′
CRY1 (F)	RN3523	8	5′-CTGTTGGTTTTGGTAGGAGAAC-3′
CRY1 (R)	RN3524	9	5′-ATTTTGCAGGGAAGCCTCTTAG-3′
NR1D1 (F)	RN3525	6	5′-CTTCCGTGACCTTTCTCAGC-3′
NR1D1 (R)	RN3526	7	5′-GCAAAGCGCACCATCAGCAC-3′
WDR5 (F)	RN3527	12-13	5′-TCTCTGTTACTGGTGGGAAGTGG-3′
WDR5 (R)	RN3528	13-14	5′-TGTTGAGATCACGACATCTGTGTG-3′
TIMELESS (F)	RN3529	12-13	5′-GCTGGCAACAGTGAATGAGATG-3′
TIMELESS (R)	RN3530	13	5′-CAGGTCACGAAGGAAAGAGC-3′
DBP (F)	RN3531	3-4	5′-GAGGAGCAGAAGGATGAGAAATAC-3′
DBP (R)	RN3532	4	5′-CGCACCGATATCTGGTTCTCC-3′
FBXL15 (F)	RN3533	3	5′-CAGAACTCCACCACCTTGACC-3′
FBXL15 (R)	RN3534	4	5′-AACGCAGCACGGGGCAGTAC-3′
ALAS1 (F)	RN3535	3	5′-GAGGATGTGCAGGAAATGAATG-3′
ALAS1 (R)	RN3536	4	5′-CCCTCCATCGGTTTTCACACT-3′
HSF1 (F)	RN3537	5	5′-CATGCCCAGCAACAGAAAGTCG-3′
HSF1 (R)	RN3538	6	5′-CTGCACCAGTGAGATCAGGAAC-3′
CHORDC1 (F)	RN3539	10-11	5′-GGGTGTGATTGATGTAAAGCG-3′
CHORDC1 (R)	RN3540	11	5′-CTGCATCGGTTCAGCTTTTCTC-3′
NPAS2 (F)	RN3541	3	5′-CCCCAAGAGTTACCTGTCCC-3′
NPAS2 (R)	RN3542	4	5′-AGGCAGGAGCTGCTGTGTGAG-3′
ARNTL (F)	RN3543	12	5′-GTGAACGGGGAAATCAGGGTG-3′
ARNTL (R)	RN3544	12-13	5′-TAGCTGTTGCCCTCTGGTCTAC-3′
ARNTL2 (F)	RN3545	9	5′-ATGGATGCTTACCCAACTCAAAG-3′
ARNTL2 R	RN3546	10	5′-CCAGCTTCTCAAGTAACCAGTG-3′
CRY2 (F)	RN3547	8	5′-GTTCTTCCACTGCTACTGCCC-3′
CRY2 (R)	RN3548	8-9	5′-AGGTATCGCCTGATGTAGTCCC-3′
PER2 (F)	RN3549	10	5′-CCTCTCCTGGGCTACCTACC-3′
PER2 (R)	RN3550	10-11	5′-GCAGGATCTTTTTGTGGATGGC-3′
PER3 (F)	RN3551	15	5′-ATAAGCCAATGCGGTTACAGC-3′
PER3 (R)	RN3552	16	5′-GTCCAGGGCTCACAGAAGAG-3′
BHLHB2 (F)	RN3553	4	5′-GTACCTGAATAAGAAGACAACTC-3′
BHLHB2 (R)	RN3554	4-5	5′-CACATCCAAAGTTCCTTCTGTAC-3′
BHLHB3 (F)	RN3555	1-2	5′-AGATTTTATAGGACTGGACTATTC-3′

Table 3.1. (continued)

Target gene (primer)[a]	Primer number	Exon	Primer sequences
BHLHB3 (R)	RN3556	2	5′-TGTCGTCTCGTTTCATGCTCC -3′
NONO (F)	RN3557	5	5′-AAGAGGGACTTCCAGAGAAGCTG-3′
NONO (R)	RN3558	5-6	5′-GGGTGGCTGCTCTCGTTCCTTG-3′
CSNK1E (F)	RN3559	7	5′-GCACATCGAGAGCAAGTTCTAC-3′
CSNK1E (R)	RN3560	8	5′-GTTGTAGTCGCCCTCAGCTC -3′
FBXL3 (F)	RN3561	3	5′-AACACTTGGACTTATTTCAACTGC-3′
FBXL3 (R)	RN3562	4	5′-GAGTTTACGAACACAACTGTCAG-3′
CIPC (F)	RN3563	2	5′-CTTGGGATGGCTGCTGCTGAG-3′
CIPC (R)	RN3564	3	5′-GGTGTTGAGGAAGGGTCTGAG-3′
EZH2 (F)	RN3565	10	5′-AAACGGGGGGAGAGAACAATG-3′
EZH2 (R)	RN3566	10-11	5′-ACCGAGAATTTGCTTCAGAGG-3′
PPARGC1A (F)	RN3567	11-12	5′-AGTAAATCTGCGGGATGATGGAG-3′
PPARGC1A (R)	RN3568	12	5′-TTCAAGAGCAGCAAAAGCATCAC-3′

[a] Abbreviations: F: forward PCR primer; R: reverse PCR primer. Gene sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/Homo_sapiens/Info/Index).

3.2.3 Chromatin immunoprecipitation (ChIP)

Approximately 1×10^6 SK-N-SH cells were plated on 35 cm^2 plates with regular media [Alpha modified MEM + 10% Fetal Bovine Serum (FBS) + 1% penicillin-streptomycin (P/S) + 1% of 100 mM L-glutamine + 1% nonessential amino acid + 1% sodium pyruvate]. When cells reached 70-80% confluence, formaldehyde was added to the media to make a final concentration of 1% and the plates were incubated at 37°C for 10 minutes to cross-link proteins and DNA. After washing with 5 ml ice-cold PBS, cells were scraped from the plates in 5 ml ice-cold PBS using a cell lifter, put into a 15 ml tube, centrifuged, re-suspended in SDS lysis buffer (ChIP assay kit, Millipore) and sonicated to shear DNA. Samples were precleared with protein G-agarose/salmon sperm beads (Millipore). Protein-DNA complexes were immunoprecipitated with anti-NRF1 (from Dr. Daniel Raines), anti-H3K4me1 [Histone H3 (mono methyl K4) antibody, Abcam, Catalog #: ab8895], anti-H3K4me3 [Anti-trimethyl-Histone H3 (Lys4), Millipore, Catalog #: 07-473], or anti-H3K9me2 [Anti-dimethyl-Histone H3 (Lys9), Millipore, Catalog #: 07-521]. Complexes were collected with Protein G agarose/salmon sperm beads and washed. Protein-DNA complexes were eluted off the beads and cross-links were reversed by incubation with NaCl at 65ºC overnight. On the next day, DNA was recovered by phenol-chloroform extraction and precipitated by ethanol. PCR was performed using the recovered immunoprecipitated DNA as templates and primers from **Table 3.2**. The PCR products were run on an agarose gel to examine the results.

Table 3.2. PCR primers for ChIP assays of circadian gene *cis*-regulatory elements, in human SK-N-SH cells or mouse Neuro2a cells.

Target gene (primer)[a]	Primer number	Primer sequences
CLOCK (F)	RN3377	5′-GCCTGCAGAGCCAGATTTCG-3′
CLOCK (R)	RN3378	5′-AAGCCGAGTCCGTGATTGCG-3′
CLOCK intron 1 (F)	RN3379	5′-GAGTGCCGGTTGGCTTCTTG-3′
CLOCK intron 1 (R)	RN3380	5′-CACTACCCCGCCTGTGTCCAG-3′
PER1 (F)	RN3381	5′-CAGCTGACGTCGGTTTCCCTG-3′
PER1 (R)	RN3382	5′-GTAGACAAAAACCTGGACACTTCC-3′
CRY1 (F)	RN3383	5′-CTGAAGGAAACCGGACAATTTC-3′
CRY1 (R)	RN3384	5′-GCTGAGACCCGGATGAGCAC-3′
NR1D1 (F)	RN3387	5′-ATCACATGGTACCTGCTCCAGTG-3′
NR1D1 (R)	RN3388	5′-GAAGTAAGTAGGTGATGGGGAGAA-3′
WDR5 (F)	RN3389	5′-CGTAGCGCTCCTCCGAGAG-3′
WDR5 (R)	RN3390	5′-GGACTCAGCTCGACAAGGC-3′
TIMELESS intron 1 (F)	RN3391	5′-GAGTGAGTGTGTGGCGAGAGG-3′
TIMELESS intron 1 (R)	RN3392	5′-GGAGACTAAGGAGCAGAGTACAG-3′
DBP intron 2 (F)	RN3393	5′-CCCTTCCCAGCGGCACATTCC -3′
DBP intron 2 (R)	RN3394	5′-CAGCTAAGGACACAGGTTCAGG -3′
FBXL15 (F)	RN3395	5′-GTCTGTTCGGGCAGTCAGG-3′
FBXL15 (R)	RN3396	5′-CAGACCCCTGCAAAGATAGAG-3′
ALAS1 (F)	RN3397	5′-ATATCCGCAGAGCCCAAGAAG-3′
ALAS1 (R)	RN3398	5′-CACTCAAGTCGAGAAGTCCAAAC-3′
CHORDC1 (F)	RN3401	5′-CTGTTGTCTGGGCTCCACTTC-3′
CHORDC1 (R)	RN3402	5′-ACGATCCGTTGCGTTTCAGG-3′
CIPC (F)	RN3403	5′-GACGACGAGGTTGTCATGGTG-3′
CIPC (R)	RN3404	5′-CTCCCCTACTTCCACCTCCTTTC-3′
EZH2 (F)	RN3405	5′-GCGTCCCTTACAGCGAACCC-3′
EZH2 (R)	RN3406	5′-CAATCGCCATCGCTTTTATTTG-3′
HSF (F)	RN3411	5′-GCACAGCCGCTTCCGACAG-3′
HSF (R)	RN3412	5′-GCGCGTGTTGGTCCCACC-3′
TEF (F)	RN3598	5′-CGCCAATCAGGGGACACAG-3′
TEF (R)	RN3599	5′-CGCCTCCTTCCATCGCTAC-3′
PPP5C (F)	RN3600	5′-TCTCATTCCCTATTGGCTCTC-3′
PPP5C (R)	RN3601	5′-GCCGCACAAGTGTCGTAAAG-3′
BTRC (F)	RN3602	5′-CGTAGCCTCAGTTTTTTTCCTG-3′
BTRC (R)	RN3603	5′-CGCCCTCTCTTACCTCTCAG-3′

[a] Abbreviations: F: forward PCR primer; R: reverse PCR primer. Gene sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/Homo_sapiens/Info/Index

and http://www.ensembl.org/Mus_musculus/Info/Index). Unless indicated otherwise, all PCR primers are located in the promoter of the indicated gene.

3.2.4 *NRF1* siRNA assays

NRF1 siRNA primers were designed (Brummelkamp *et al*., 2002; Smith *et al*., 2004), cloned into the pSUPER siRNA vector and then sequenced to confirm the correct structure. SK-N- SH cells were sub-cultured onto 75 cm^2 flasks 1 day before the nucleofection and reached 80-90% confluence. On the day of nucleofection, 4 μg of pSUPER-NRF1 DNA were transfected into 1×10^6 SK-N-SH cells using an Amaxa cuvette (Lonza) with Nucleofector® (Lonza) and the transfected cells were plated onto a 6-well plate. 24 hr after nucleofection, RNA was extracted with Trizol and treated by DNase I to remove genomic DNA. 1 μg of RNA from each sample was reverse transcribed into cDNA using SuperScript® III reverse transcriptase (Invitrogen) according to the manufacture's instructions. RT-PCR was performed using primers for each gene (**Table 3.1**) to examine gene expression with *GAPDH* as a negative control and *NRF1* as a positive control to ensure effective down-regulation by siRNA.

3.2.5 Reporter constructs and luciferase reporter assays

Based on sequence alignments of conserved promoter sequence of *CLOCK*, *PER1*, *CRY1*, *NR1D1* and *DBP*, and enhancer sequence of *PER1* and *DBP*, luciferase reporter constructs were made by PCR cloning using primers in **Table 3.3** to clone the minimal promoter (all five genes) and enhancer (*PER1* intron 1 and *DBP* intron 2) sequences using genomic DNA from human SK-N-SH cells as templates. PCR products were digested with *Bgl*II and *Hin*dIII for promoters or

*Bam*HI and *Sal*I for enhancers, and ligated products directionally into the pGL3basic vector (Promega). For the *CLOCK* promoter, *NR1D1* promoter, and *DBP* enhancer, site-directed mutations in the putative NRF1 binding sites were performed using primers in **Table 3.3** with a GeneTailor™ Site-Directed Mutagenesis Kit (Invitrogen) according to the manufacturer's instruction and the mutated products were amplified by Platinum *Taq* DNA polymerase High Fidelity (Invitrogen). However, by sequence analysis, we found four unexpected mutations, which were introduced by PCR, in either the promoter or in the luciferase gene of the *DBP*-pr+enh-NRF1 construct, and one mutation in the luciferase gene of the *CLOCK*-pr-mu construct. To correct these errors, *Bam*HI and *Sal*I were used to isolate the desired *DBP* enhancer fragment carrying the NRF1 site mutation, while *Bgl*II and *Hin*dIII were used to isolate the desired *CLOCK* promoter fragment with the NRF1 site mutation, and each used to generate the expected constructs, respectively. Both new constructs were sequenced to confirm correct insertions. To make the NRF1 site mutation in the putative *PER1* enhancer, a minigene was designed (**Table 3.3**) and synthesized by Integrated DNA Technologies, Inc. The mutated *PER1* enhancer fragment was released from the synthetic construct and ligated into the *PER1*-pr following *Bam*HI and *Sal*I digestion. Subsequently, five groups of constructs were transfected into SK-N-SH human neuroblastoma cells in a 6-well plate or NIH3T3 cells in a 24-well plate using Lipofectamine 2000 (Invitrogen). For each well, 1 μg of pGL3basic-derived vector containing a firefly luciferase gene was co-transfected with 25 ng of pRL-SV40 vector containing a *Renilla* luciferase gene. Cells were lysed 24 hr after transfection by passive lysis buffer (Promega). Luciferase activity was measured using a Dual-luciferase reporter assay (Promega) on a Luminometer $20/20^n$ device (Turner BioSystems). For each sample, the firefly luciferase activity was normalized to *Renilla* luciferase activity. The relative luciferase activity was normalized to pGL3basic vector luciferase

activities. Each experiment was repeated three times to collect data for statistical analysis and analyzed by a two-tailed t-test.

Table 3.3. PCR primers and a minigene used for luciferase constructs and mutagenesis.

Target gene (primer or oligonucleotide)[a]	Primer number	Primer sequences
CLOCK promoter (F)	RN3644	5′-GACT**AGATCT**GCGGCTCCGTGCTGCCTAAC-3′
CLOCK promoter (R)	RN3645	5′-GAAA**AAGCTT**CAAGCCGAGTCCGTGATTG-3′
DBP promoter (F)	RN3647	5′-CACA**AGATCT**ATCAGGCAGCACGAGCAGAGC-3′
DBP promoter (R)	RN3648	5′-GGTT**AAGCTT**CCAAAGCAAACTTCTTTCGC-3′
DBP enhancer (F1)	RN3649	5′-CTTT**GGATCC**CGGCACATTCCTGCGCCACG-3′
DBP enhancer (F2)	RN3650	5′-CAAT**GTCGAC**CTAAGGACACAGGTTCAGG-3′
DBP enhancer (R1)	RN3651	5′-CTTT**GGATCC**TGAGTCCTCCCATCCCTCCTC-3′
DBP enhancer (R2)	RN3652	5′-CAAT**GTCGAC**TGCTGACGCCTGCTTCCCTC-3′
NR1D1 promoter (F)	RN3653	5′-CTTT**AGATCT**CGCAGTCCGCCCACTTTGTC-3′
NR1D1 promoter (R)	RN3654	5′-CTTT**AAGCTT**CTTTTGCCCGAGCCTTTCCCTG-3′
PER1 promoter (F)	RN3655	5′-CACA**AGATCT**CTCTTCAGCCCAGCACCAGC-3′
PER1 promoter (R)	RN3656	5′-CTTT**AAGCTT**TCCACCGGGCGGGCGGGGAG-3′
PER1 enhancer (F)	RN3657	5′-TTAA**GGATCC**CGTTCCCCCACTTCCGCCG-3′
PER1 enhancer (R)	RN3658	5′-CACA**GTCGAC**CTCCAGGCTAATTTTATC-3′
CRY1 promoter (F1)	RN3659	5′-CTTT**AGATCT**CTTACCCTCTGGAACGCAGC-3′
CRY1 promoter (F2)	RN3675	5′-CTTT**AGATCT**CAAGCTCCCCGCATCCCC-3′
CRY1 promoter (R1)	RN3660	5′-CCAT**AAGCTT**TGTTTACTACACTGGCTCGG-3′
CRY1 promoter (R2)	RN3661	5′-CCCT**AAGCTT**CCTCGGCCCCGCCCCCG-3′
CLOCK promoter mutation (F)	RN3693	5′-GTGCTGCCTAACGGGGCAAGT***aaa***AT***aaa***CACCGAGCCG-3′
CLOCK promoter mutation (R)	RN3852	5′-ACTTGCCCCGTTAGGCAGCACGGAGCCGCGC-3′
DBP enhancer mutation (F)	RN3676	5′-CTCCCATCCCTCCTCCGCGCTTCGC***aa***AG***aaa***CCTTCAAGCG-3′
DBP enhancer mutation (R)	RN3677	5′-GCGAAGCGCGGAGGAGGGATGGGAGGACTCACGTGG-3′
NR1D1 promoter mutation (F)	RN3855	5′-TCCCGACAGTCTTGTCGTTG***aaa***AG***aaa***CGCAAGAGCTC-3′
NR1D1 promoter mutation (R)	RN3856	5′-CAACGACAAGACTGTCGGGATTTGTAGTCCACC-3′

Table 3.3. (continued)

Target gene (primer or oligonucleotide)[a]	Primer number	Primer sequences
PER1 enhancer mutation minigene (sense-strand only is shown)	RN3717	5′- TTAA**GGATCC**CGTTCCCCCACTTCCGCCGGGAAATGG GGGAGGGGTCGCTCCTCCCGCCCTCCTGTGGTCCCTC CAGCAACCGCTGAGCTCAGCAGCTGACGTCGGTTTCC CTGGCGACCGTGGCGGCGGCGGAAGCGCGTGGTGGGG CCGCGCACGTCGGCG***aaa***AT***aa***GCAGCGGGGGTGGCA CCGCCCCCGGATAAAATTAGCCTGGAG**GTCGAC**TGTG -3′

[a] Abbreviations: F: forward PCR primer; R: reverse PCR primer. Gene sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/Homo_sapiens/Info/Index). For primer sequences, bold and underline represent restriction enzyme sites used for cloning purposes, while bold, italicized lowercase letters represent mutations (shown only for the sense strand of oligonucleotides). Each of the five mutations generated is within an NRF1 site within the corresponding wildtype (WT) DNA sequence.

3.2.6 Luciferase reporter assays on co-transfection of the *DBP* promoter-enhancer luciferase construct and expression vectors for CLOCK, BMAL1 and NRF1

5×10^4 NIH3T3 cells were plated into a 24-well plate a day before the transfection. On the day of the experiment, each well was transfected with 250 ng pGL3b-DBP pr+enh construct, 25 ng pRL vector and/or (1) 125 ng pcDNA3-HA-Clock and 125 ng pcDNA3-HA-BMAL1 (these expression vectors were a gift from Dr. M.P. Antoch, Roswell Park Cancer Institute, Buffalo; Kondratov *et al*., 2006); (2) 125 ng pcDNA3-NRF1-VP16 (a gift from Dr. T. Gulick; Ramachandran *et al*., 2008); and (3) 125 ng of each of the three expression vectors, using

Lipofectamine 2000 for 20 minutes, the medium was changed to 500 µl/well OPTI-Medium. 24 hr after the transfection, cells were lysed by passive lysis buffer (Promega) and luciferase activity was measured by a Luminometer $20/20^n$ device (Turner BioSystems). For each sample, the firefly luciferase activity was normalized to *Renilla* luciferase activity. The relative luciferase activity was normalized to the luciferase activity of the pGL3basic vector. Each experiment was repeated three times to collect data for statistical analysis and analyzed by a two-tailed t-test.

3.2.7 Co-immunoprecipitation

Approximately 2×10^6 NIH3T3 cells were plated in 35 cm^2 plates with regular medium [(Dulbecco's modified Eagle medium (DMEM)+10% calf bovine serum (CBS) + 1% penicillin-streptomycin (P/S)]. When cells reached 80% confluence, media was removed and cells were washed once with 5 ml ice-cold PBS/Phosphatase Inhibitors (Nuclear Complex Co-IP Kit, Active Motif). By adding another 5 ml ice-cold PBS/Phosphatase Inhibitors, cells were removed gently by a cell lifter and collected by centrifugation, and cell pellets were re-suspended in 1 ml 1× Hypotonic Buffer with Detergent and the cells lysed by incubation on ice for 15 min. After centrifugation, pellets (nuclear fraction) were re-suspended in 200 µl Complete Digestion Buffer with 1 µl Enzymatic Shearing Cocktails and incubated at 4°C for 90 min. To stop the reaction, 4 µl 0.5 M EDTA was added to each sample and centrifuged at 14,000 × g at 4°C for 10 min to obtain supernatant. Protein concentration was measured by Bovine Gamma Globulin (BGG) Protein Standards assays (Pierce). Control agarose resin (Cross-linked Co-IP kit, Pierce) was used to pre-clear the cell lysates before immunoprecipitation.

Anti-NRF1 (received from Dr. Daniel Reines) and anti-BMAL1 (N-20, Santa Cruz Biotechnology) were each cross-linked to Protein A/G Plus Agarose on Pierce Spin Column by

DSS solution (Cross-linked Co-IP kit, Pierce). For each co-IP reaction, 500 μg of nuclear extracts were diluted into a total of 600 μl IP Incubation Buffer and incubated overnight at 4°C with rotation. On the next day, the columns were washed and co-IP products were collected by elution buffer according to the manufacturer's instruction. The antigens were then used for western blot analysis. Cell lysates incubated with control agarose resin overnight were used as a control according to the manufacturer's instruction.

3.2.8 Western blot

Immunoprecipitated protein samples were loaded and separated on 4-12% Bis-Tris gels (BioRad, Criterion XT Precast Gel) and transferred onto a nitrocellulose membrane. IP samples by anti-rabbit NRF1 and anti-rabbit BMAL1 were detected on a western blot with anti-CLOCK (S-19, Santa Cruz Biotechnology) as a primary antibody and horseradish peroxidase (HRP) conjugated donkey anti-goat secondary antibody (Santa Cruz Biotechnology) as a secondary antibody. The membrane was detected by chemiluminescence (Western Blotting Luminol Reagent, Santa Cruz Biotechnology) using BioMax film.

3.2.9 Cell cycling studies

Using NIH3T3 cells, 1×10^6 cells were plated on 35 cm^2 plates or 5×10^4 cells were plated on 6-well plates with regular medium [(Dulbecco's modified Eagle medium (DMEM) + 10% calf bovine serum (CBS) + 1% penicillin-streptomycin (P/S)]. After two days, when cells reach 80% confluence, cells were changed into 1% serum medium (DMEM + 1% BCS + 1% P/S) for 1 day. At t=0, cells were replaced with serum-rich medium (50% horse serum + DMEM + 1%

P/S, Balsalobre *et al*., 1998) for two hours. After that, cells were changed into serum-free medium (DMEM + 1% P/S). At each indicated time point, cells were washed by PBS, trypsinized, collected by centrifugation and frozen at -80°C for protein analyses or Trizol (Invitrogen) applied directly onto the cells, collected and frozen at -80°C for RNA analyses.

RNAs was extracted according to the manufacturer's instruction (miRNA kit, Qiagen) and treated by RNase-free DNase I (Promega) to remove genomic DNA. 1 μg of RNA sample from each time point was reverse transcribed by SuperScript® III reverse transcriptase (Invitrogen) into cDNA using hexamer oligonucleotides and oligo dT primers in a 20 μl reaction. Expression of *Dbp, Nr1d1,* and *Nrf1* genes using the respective primer sets (**Table 3.4**) was analyzed by quantitative real-time PCR (QPCR) using an Applied Biosystems 7300 Real Time PCR system. SYBR green (Power SYBR® Green PCR Master Mix, Applied Biosystems) was used as a fluorescent dye and each cDNA sample was loaded in triplicate. For each sample, a ΔC_T value was obtained by normalizing each gene of interest against *GAPDH* in each assay. The $\Delta\Delta C_T$ value was obtained by subtracting the ΔC_T values for each sample from the ΔC_T values of *GAPDH* in each corresponding assay to calculate a relative $\Delta\Delta C_T$. All $\Delta\Delta C_T$ values were converted to fold change.

For protein analysis, nuclear extracts were obtained from each time point (Nuclear Extract Kit, Active Motif) and protein concentration was measured. Samples were loaded onto a Criterion XT 4-12% Bis-Tris Gel (Bio-Rad) for protein separation, followed by transfer to a western blot and subsequent detection using anti-rabbit NRF1 (received from Dr. Daniel Reines), anti-rabbit DBP (Santa Cruz Biotechnology), or anti-rabbit β-actin (ab8227, Abcam).

Table 3.4. RT-PCR primers for gene expression studies (mouse NIH3T3 cells).

Target gene (primer)[a]	Primer number	Exon	Primer sequences
Clock (F)	RN3770	7	5′-CTTAGTAATGAAGAGTTTACACAG-3′
Clock (R)	RN3771	8	5′-TAACGAAGTTACACTCTCAGATAC-3′
Per1 (F)	RN3772	16-17	5′-CCAAGAAAGATCCGTCGTCAGC-3′
Per1 (R)	RN3773	17	5′-AGGGCGAGCGGGCTCAGGGT-3′
Cry1 (F)	RN3774	8	5′-CTGTGGGTTTTGGTAGGAGGAC-3′
Cry1 (R)	RN3524	9	5′-ATTTTGCAGGGAAGCCTCTTAG-3′
Nr1d1 (F)	RN3525	6	5′-CTTCCGTGACCTTTCTCAGC-3′
Nr1d11 (R)	RN3526	7	5′-GCAAAGCGCACCATCAGCAC-3′
Dbp (F)	RN3775	3-4	5′-CTGAGGAACAGAAGGATGAGAAG-3′
Dbp (R)	RN3776	4	5′-ATCTGGTTCTCCTTGAGTCTTC-3′
Nrf1 (F)	RN3807	9	5′-ATTACTCTGCTGTGGCTGATGG-3′
Nrf1 (R)	RN3808	10	5′-CGTCGTCTGGATGGTCATTTC-3′
Nrf1 3’-UTR (F)	RN3945	12	5'-GCCCGTGTTCCTTTGTGGTG-3'
Nrf1 3’-UTR (R)	RN3946	12	5'-GGAGAACAACACAGATTCCATGC-3′
Arntl (F)	RN3779	12-13	5′-TACCTGTTCAAAGAAAAAAGCAG-3′
Arntl (R)	RN3780	13	5′-GGCTCATTGTCTTCGTCCAGC-3′
Per2 (F)	RN3777	10	5′-CCTCCTGGGCTATCTACCTC-3′
Per2 (R)	RN3778	11	5′-CGGTGGACAGCCTTTCGATTA-3′
Clock alternative splicing (F)	RN4066	15	5′-GCGAGAACTTGGCATTGAAGAG-3′
Clock alternative splicing (R)	RN4067	21	5′-CTGTGTCCACTCATTACACTCTGTTG-3′

[a] Abbreviations: F: forward PCR primer; R: reverse PCR primer; ex: exon; in: intron. Gene sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/Mus_musculus/Info/Index).

3.3 RESULTS

3.3.1 Putative NRF1 sites in 5' regulatory elements of ~56% circadian regulatory genes

To determine whether NRF1 may have a role in regulating circadian genes, we first identified 25 potential NRF1 target genes (**Table 3.5**) from bioinformatics searches of 45 candidate genes (**Table 3.5 & Table 3.6**) in human and mouse genome sequences, with strong matches to the NRF1 canonical binding site, 5'-YGCGCANGCGCR-3'. This sequence is a pyrimidine (Y)-purine (R) repeating motif, with both a direct repeat and a palindrome (in each case, 5'-YGCGCR-3'); either of the latter two features is likely responsible for NRF1 binding DNA as a homodimer (Gugneja & Scarpulla, 1997; Gomez-Cuadrado *et al*., 1995). Prior biochemical footprinting studies have shown that NRF1 contacts most or all guanines on both strands of DNA (Virbasius & Scarpulla, 1994), and extensive phylogenetics, *in vitro* and *in vivo* studies in our laboratory and others (Rodriguez-Jato *et al*., 2005, unpublished data) are consistent with the essential nature of the two GCGC sub-motifs. One nucleotide substitution in one GCGC motif is allowed, but reduces affinity, while two substitutions prevent NRF1 binding. Only the Y, N, and R positions are variable, with consensus sequences at these positions increasing affinity. Second, for candidate NRF1 target circadian genes in human and mouse (**Table 3.5**), we assessed phylogenetic conservation by multi-sequence alignment for which the majority of identified TF sites were present in all or most sequenced mammalian genomes (**Figure 3.1 & Appendix A**), indicating functional importance.

(a) Multi-sequence alignment of *CLOCK* promoter

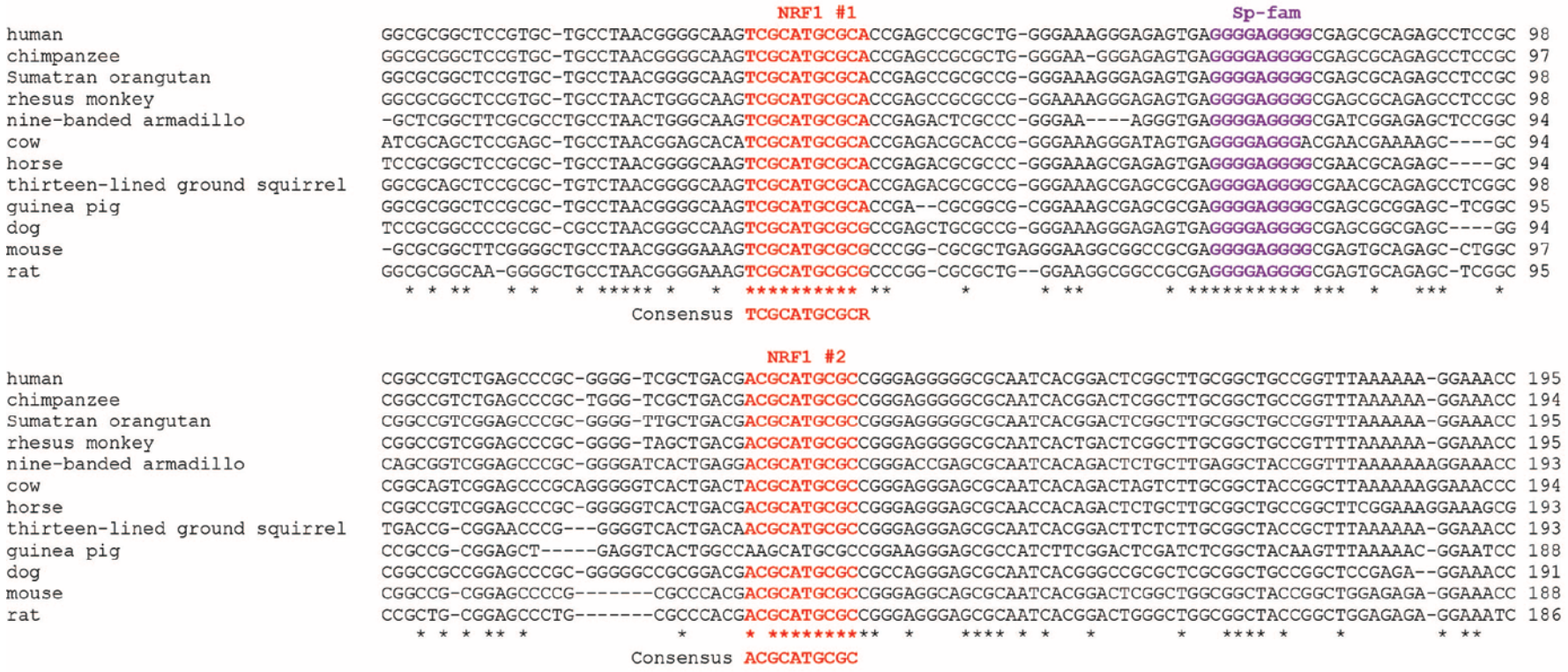


Figure 3.1. (a) (continued below)

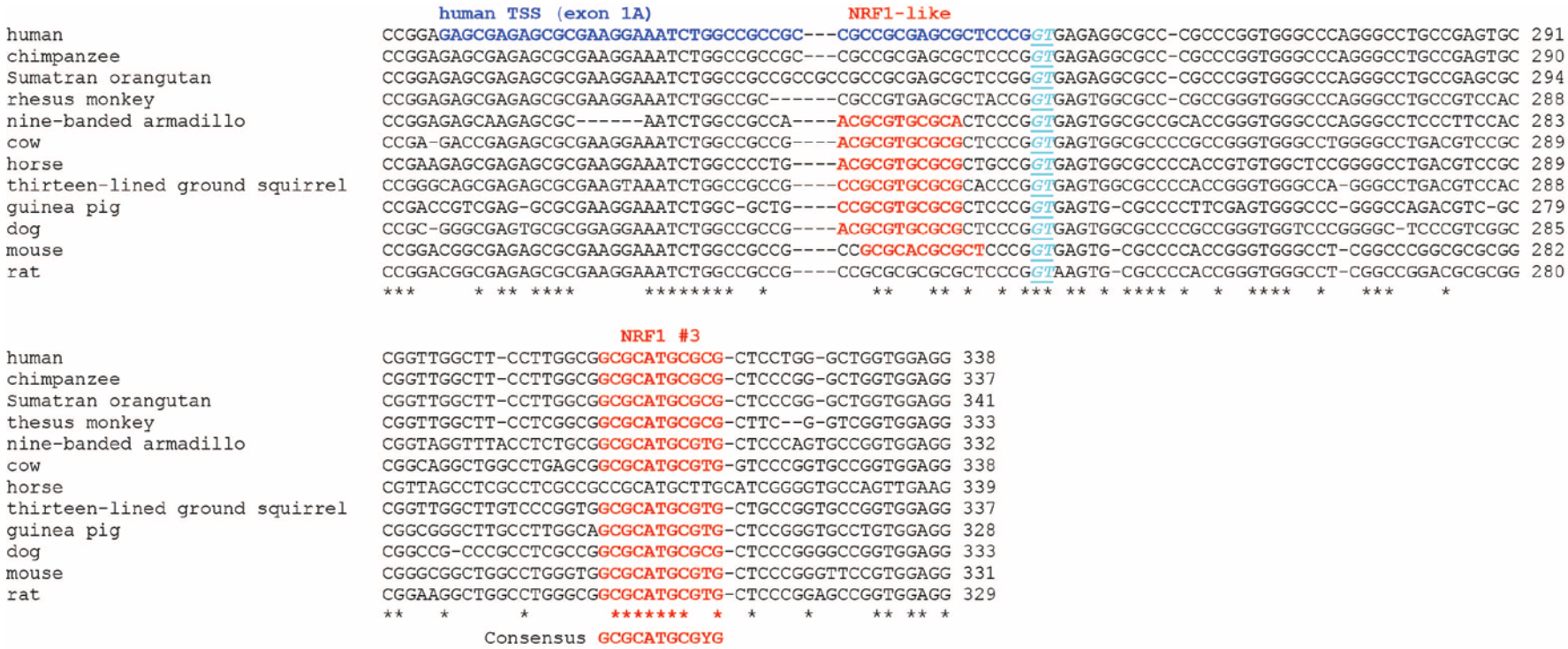


Figure 3.1. (a) (continued below)

(b) Multi-sequence alignment of *CRY1* promoter

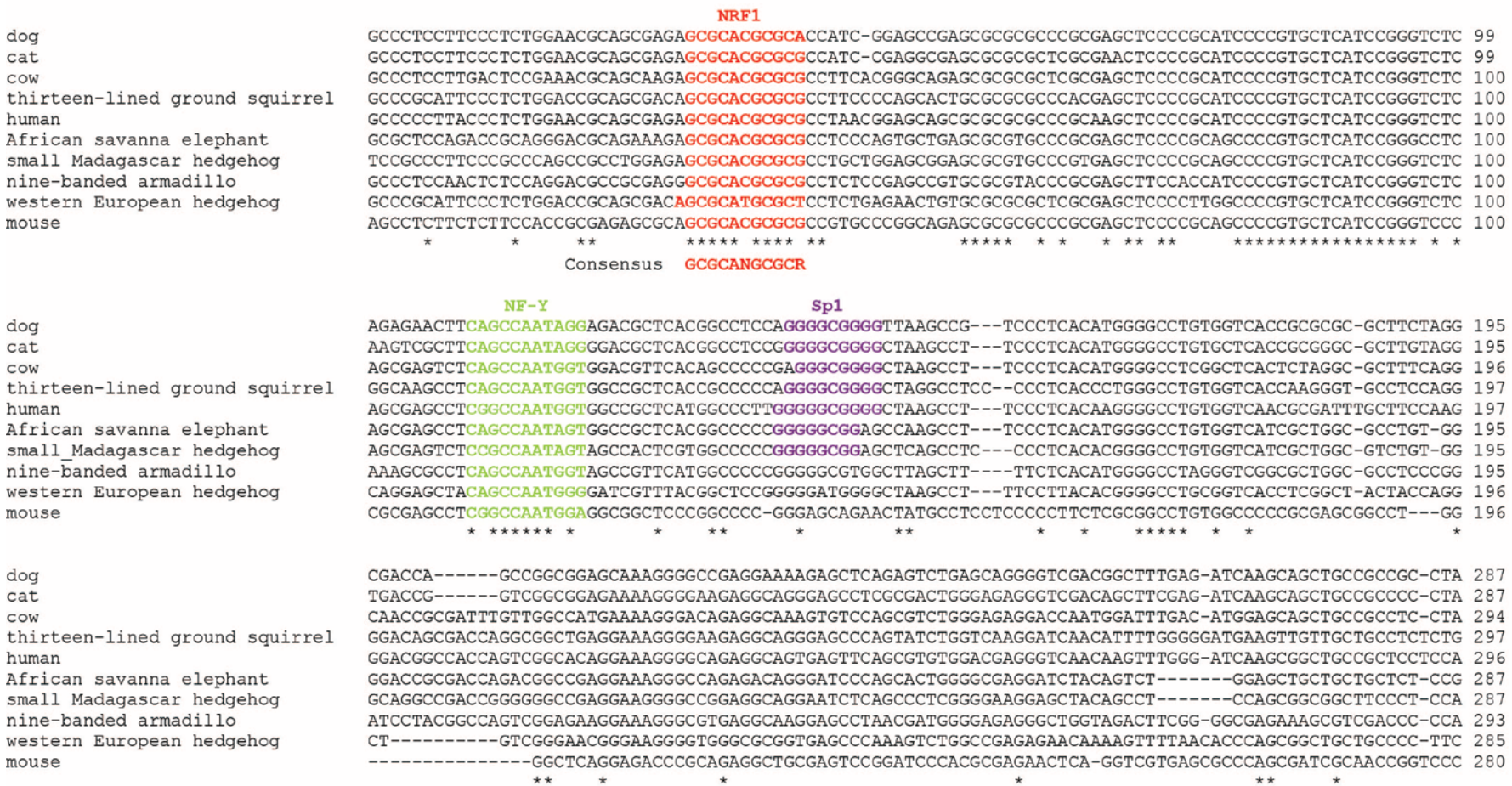


Figure 3.1. (b) (continued below)

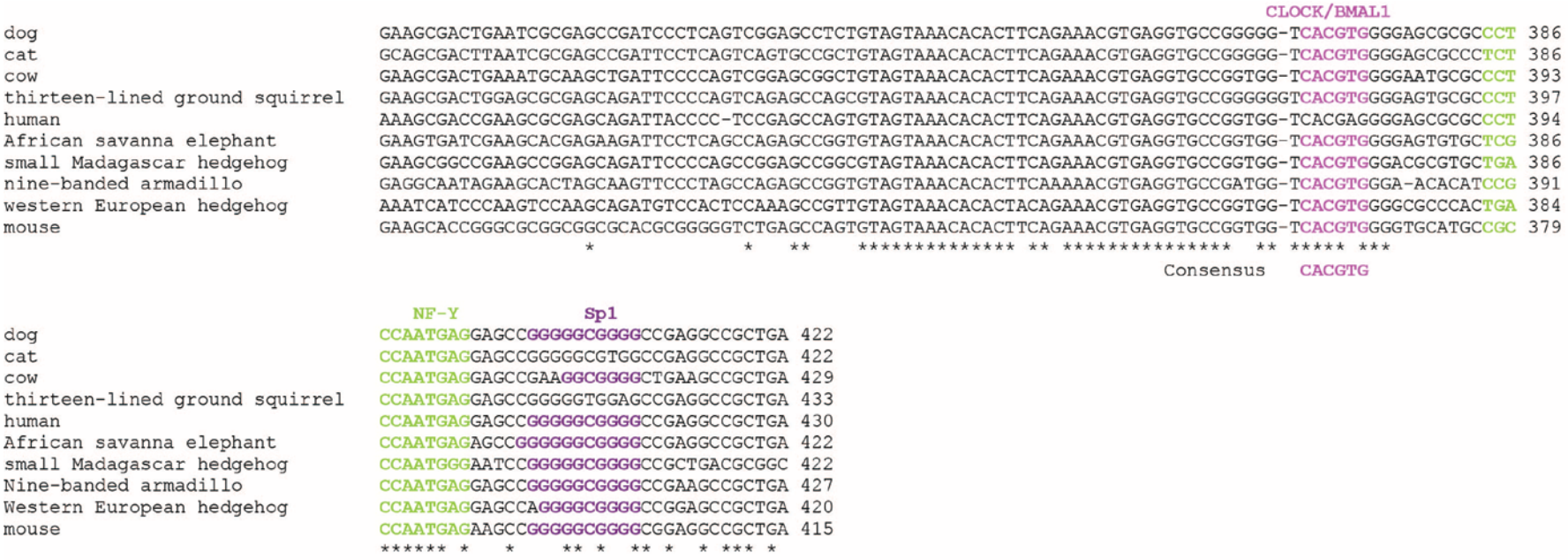


Figure 3.1. (b) (continued below)

(c) Multi-sequence alignment of *PER1* promoter

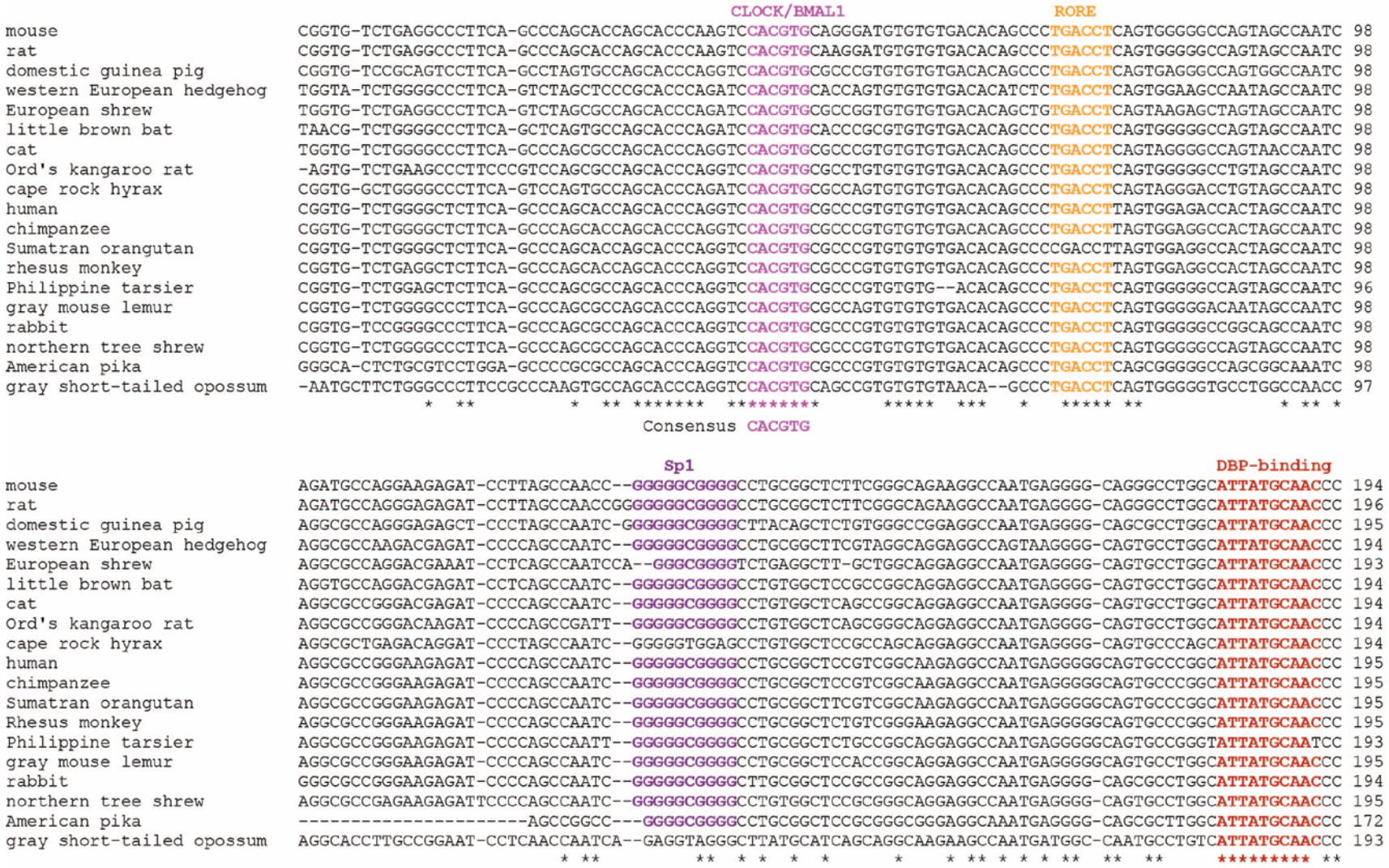


Figure 3.1. (c) (continued below)

(d) Multi-sequence alignment of *PER1* intron 1

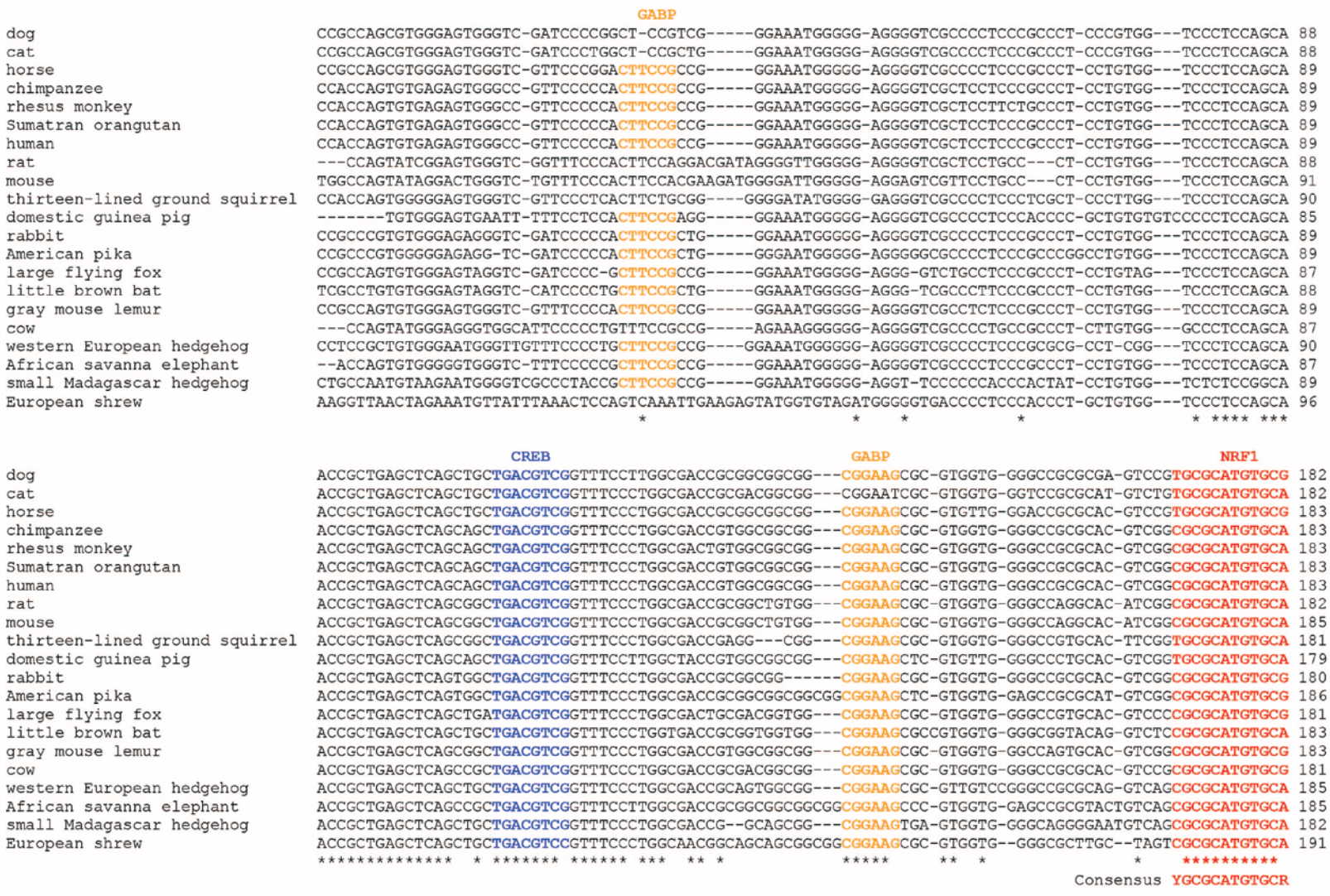


Figure 3.1. (d) (continued below)

Figure 3.1. (d) (continued below)

(e) Multi-sequence alignment of *NR1D1* promoter

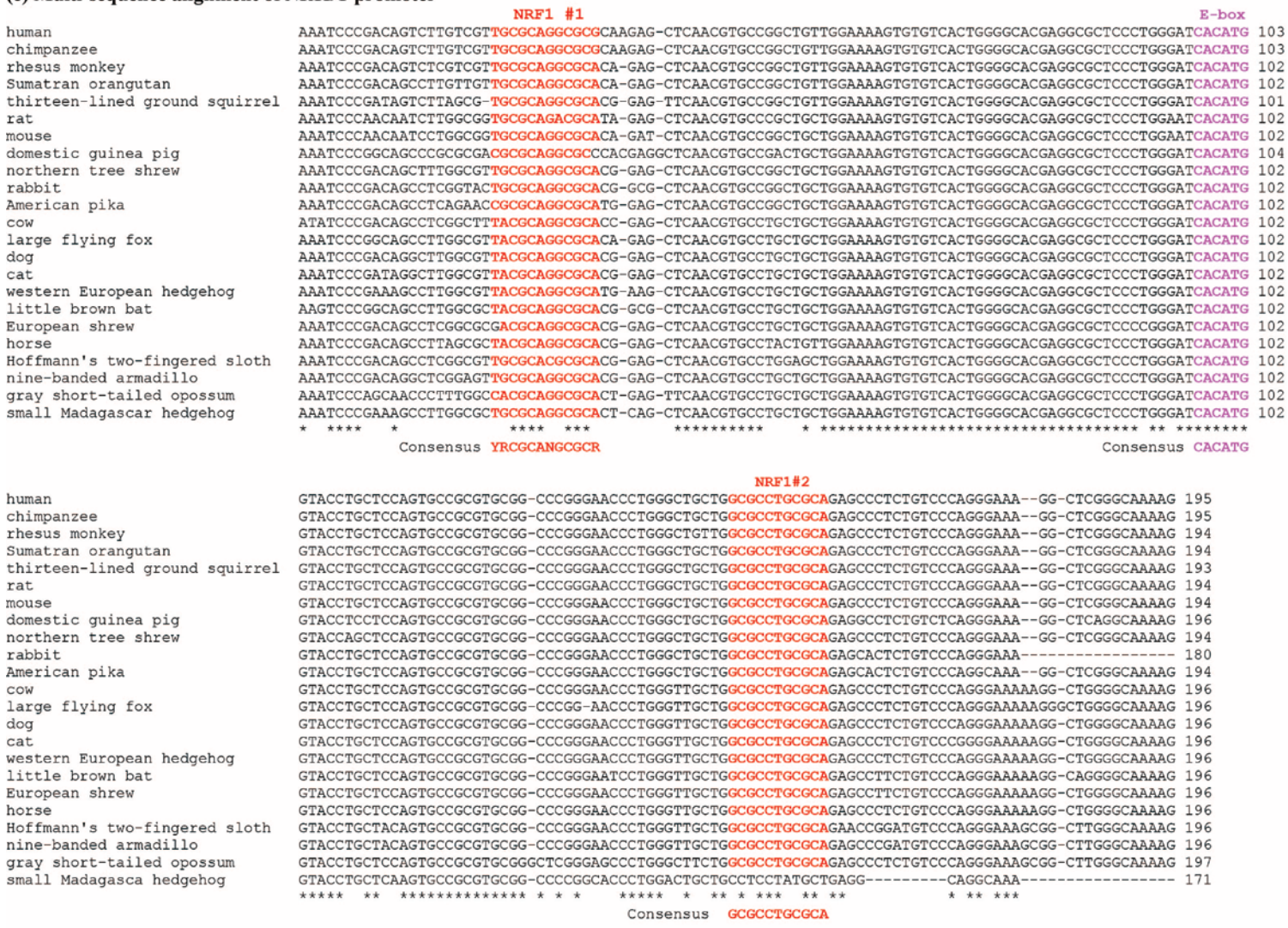


Figure 3.1. (e) (continued below)

(f) Multi-sequence alignment of *DBP* promoter

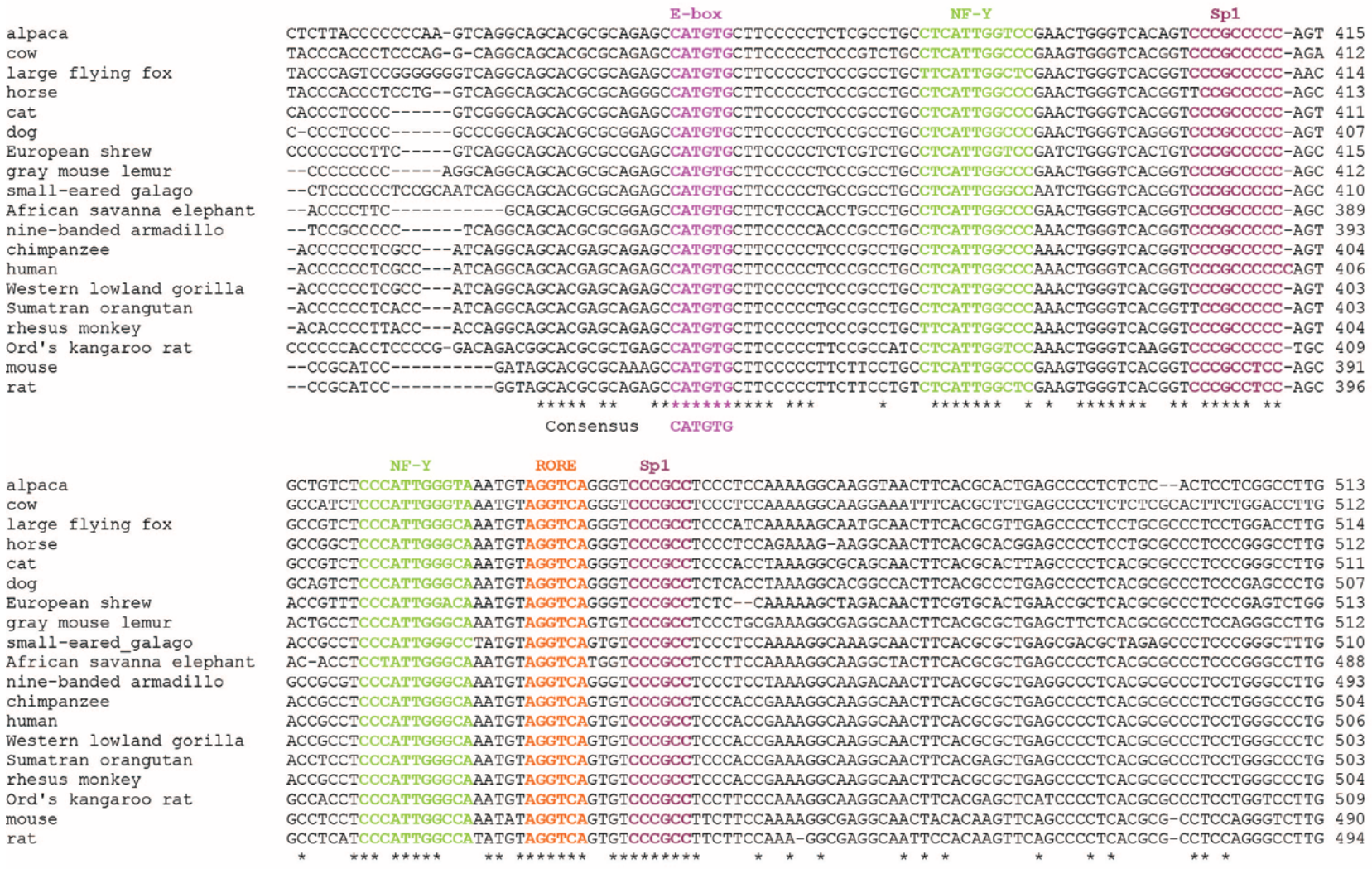


Figure 3.1. (f) (continued below)

Figure 3.1. (f) (continued below)

(g) Multi-sequence alignment of *DBP* intron 2 enhancer

Figure 3.1. Multi-sequence alignments for (a) *CLOCK* promoter; (b) *CRY1* promoter; (c) *PER1* promoter; (d) *PER1* intron 1; (e) *NR1D1* promoter; (f) *DBP* promoter; and (g) *DBP* intron 2 enhancer. Consensus sequences for CLOCK/BMAL1 (E-box) and NRF1 motifs are highlighted in pink and red, respectively.

Table 3.5. NRF1 targets include 25 genes encoding core, output, or input circadian regulators.

Gene	NRF1 site	Human chromosome	Circadian regulatory function	Circadian reference
CLOCK	Promoter, Intron 1	4q12	bHLH TF; forms heterodimer with BMAL1; binds to E-box to activate gene expression; part of positive limb of circadian rhythms; possesses histone acetyltransferase activity	Steeves *et al*., 1999; Hirayama *et al*., 2007; Kondratov *et al*., 2003
PER1	Intron 1	17p13.1	Negative regulator of circadian rhythms; heterodimerizes with CRY1; *PER1* expression is activated by CLOCK/BMAL1 via binding to a promoter E-box element	Motzkus *et al*., 2007
CRY1	Promoter	12q23.3	A photoreceptor; as of *PER1*	Griffin *et al*., 1999; Etchegaray *et al*., 2003; Ueda *et al*., 2005; Langmesser *et al*., 2008
NR1D1[b]	Promoter	17q21.1	Encodes an orphan nuclear receptor; activated by CLOCK and BMAL1 and repressed by PER and CRY; a heme sensor that coordinates metabolic and circadian pathways	Preitner *et al*., 2002; Yin *et al*., 2007
DBP	Intron 2	19q13.33	A member of PAR bZIP TFs; *DBP* gene expression regulated by CLOCK/BMAL1 complex	Ripperger & Schibler U, 2006; Stratmann *et al*., 2010
BTRC[a]	Promoter	10q24.32	Part of the E3 ubiquitin ligase complex, mediates PER1 and PER2 degradation	Shirogane *et al*., 2005; Reischl *et al*., 2007;
PPP5C	Promoter	19q13.3	Part of an E3 ubiquitin ligase complex, mediates PER1 and PER2 degradation	Partch *et al*., 2006

Table 3.5. (continued)

Gene	NRF1 site	Human chromosome	Circadian regulatory function	Circadian reference
FBXL15	Promoter	10q24.32	Protein phosphatase 5; activates the kinase activity of CKIε (encoded by *CSNK1E*) by preventing autophosphorylation	Koh *et al*., 2006
WDR5	Promoter	9q34.2	A component of the histone methyltransferase complex; involved in H3K4 methylation; associates with PER1	Brown *et al*., 2005; Schuetz *et al*., 2006;
TIMELESS	Intron 1	12q13.3	Interacts with CLOCK in circadian rhythms; interacts with Chk1 and the ATR-ATRIP complex in cell cycle checkpoints	Unsal-Kacmaz *et al*., 2005
TEF	Promoter	22q13.2	PAR bZip protein; circadian output gene and TF similar to DBP	Fonjallaz *et al*., 1996
HSF1	Promoter	8q24.3	Activates expression of heat shock proteins by thermal stress; maintains physiological homeostasis of circadian rhythms	Reinke *et al*., 2008; Kim *et al*., 2005
EZH2	Promoter	7q36.1	A polycomb group enzyme; a component of CLOCK/BMAL1 complex; can bind to di- or trimethylation of H3K27 on *Per1* and *Per2* promoters; enhances mCRY-mediated transcriptional repression	Etchegaray *et al*., 2006;
CIPC	Promoter	14p24.3	Additional negative-feedback regulator to CLOCK/BMAL1	Zhao *et al*., 2007
CHORDC1	Promoter	11q14.3	A potential transcription regulator during development and has circadian expression pattern in mammalian brain	Gerstner & Landry, 2007
ALAS1	Promoter	3p21.2	Regulated by CLOCK; rate limiting enzyme in heme biosynthesis	Kaasik & Lee, 2004;

Table 3.5. (continued)

Gene	NRF1 site	Human chromosome	Circadian regulatory function	Circadian reference
FMR1	Promoter	Xq27.3	homolog *dfmr1* (*Drosophila*) mutant shows erratic locomotor activity; Fmr1 KO mice have a shorter running period in 12 hr dark-12 hr dark cycles	Dockendorff *et al*., 2002; Inoue *et al*., 2002; Zhang *et al*., 2008;
FXR2	Promoter	17p13.1	Autosomal homolog of FMR1, similar function as *FMR1*	Zhang *et al*., 2008
CSNK1E	Intron 1	22q13.1	A member of the casein kinase I (CKI) gene family; phosphorylates PER, CRY, and BMAL1	Eide *et al*., 2002; Etchegaray *et al*., 2009
NR1D2[b]	Intron 1	3p24.2	Belongs to REV-ERB family, as for *NR1D1*, uses heme as a ligand	Raghuram *et al*., 2007; Liu *et al*., 2008;
CREM	Promoter	10p11.21	a bZIP TF that binds to the cAMP responsive element, encodes the transcriptional repressor ICER, whose levels in SCN is regulated by light and expressed at high level during the night	Stehle *et al*., 1993; Foulkes, *et al*., 1996;
MYBBP1A	Promoter	17p13.3	Interacts with mCRY1, co-repressor of *PER2* promoter	Hara *et al*., 2009
RAB3A	Intron 1	19p13.2	A ras-associated binding protein; mutation of mouse gene shows abnormal circadian period	Kapfhamer *et al*., 2002
ATF5	Intron 1	19q13.3	Activating transcription factor 5; regulated by CLOCK/BMAL1	Lemos *et al*., 2007
PROK2	Promoter	3p13	A circadian output gene expressed in the SCN; predominantly entrained by endogenous clock, and light is a modulator for its rhythmic expression	Cheng *et al*., 2002; Cheng *et al*., 2005

[a] *NR1D1* is also known as *REV-ERBα*, *NR1D2* as *REV-ERBβ*, and *BTRC* as ß-TrCP1.

3.3.2 NRF1 binds to 5' regulatory elements of a subset of genes involved in circadian regulation and regulates their gene expression

To test if the predicted NRF1 targets are functional, ChIP assay was used to examine *in vivo* binding of NRF1 in human SK-N-SH neuroblastoma cells. Using antibody against NRF1 in ChIP, we showed that NRF1 bound to all predicted targets for 16 tested genes, including *CLOCK*, *PER1*, *CRY1*, *DBP* and *NR1D1* (**Figure 3.2**).

Similarly, to test the transcriptional regulatory role of NRF1, we used a pSUPER siRNA plasmid (Brummelkamp *et al*., 2002; Smith *et al*., 2004) that expresses a shRNA targeting *NRF1* exon 2 to knockdown *NRF1* mRNA levels. By knockdown of *NRF1* mRNA to ~50% levels, we found that all predicted target genes have reduced mRNA levels, including *CLOCK*, *PER1*, *CRY1*, *DBP,* and *NR1D1*, but not those for the *GAPDH* control (**Figure 3.3**). Surprisingly, we observed significant knockdown of *BMAL1* mRNA levels (**Figure 3.3**). *BMAL1* is not predicted to be a direct target of NRF1 based on the lack of a canonical NRF1 site from ~50 kb upstream of the *BMAL1* TSS, all intronic sequence, through ~50 kb downstream of the 3'-UTR. This observation indicates the possibility of a secondary effect on *BMAL1*, although an unknown mechanism by which NRF1 can regulate the circadian system cannot be excluded.

Table 3.6. 21 genes encoding core, output, or input circadian regulators that do not have promoter or intragenic canonical NRF1 sites.

Gene[a]	NRF1 site	Human chromosome	Circadian regulatory function	Reference
BMAL1	none	11p15.2	Heterodimerizes with CLOCK	Hirayama *et al*., 2007
CRY2	none	11p11.2	Paralog of CRY1	Griffin *et al*., 1999; Langmesser *et al*., 2008
PER2	none	2q37.3	Paralog of PER1	von Schantz *et al*., 2006; Kaasik & Lee, 2004
PER3	none	1p36.23	Paralog of PER1	von Schantz *et al*., 2006
NPAS2	none	2q11.2	CLOCK-related; heterodimerizes with BMAL1	Kaasik & Lee, 2004;
BMAL2	none	12p11.23	Heterodimerizes with CLOCK, paralog of BMAL1	Sasaki *et al*., 2009; Shi *et al*., 2010
DEC1	none	3p26.1	bHLH TF; interacts with ARNTL and suppresses CLOCK/BMAL1 complex	Sato *et al*., 2004; Butler *et al*., 2004; Li *et al*., 2003; Honma *et al*., 2002
DEC2	none	12p12.1	bHLH TF; interacts with BMAL1 and suppresses CLOCK/BMAL1 complex; expression is mediated by DEC1	Noshiro *et al*., 2004; Butler *et al*., 2004
FBXL3	none	13q22.3	Forms SCF ubiquitin ligase complexes with Skp1 and cullin proteins to ubiquitinate and degrade CRY1 and CRY2	Busino *et al*., 2007;
NONO	none	Xq13.1	Associates with PER; acts as a bridge between the CREB/TORC complex and RNA polymerase II	Brown *et al*., 2005; Amelio *et al*., 2007;
PPARGC1A	none	4p15.2	A transcriptional coactivator regulates genes involved in energy metabolism; stimulates BMAL1, NR1D1, NRF1 and NRF2 (GABP)	Liu *et al*., 2007

Table 3.6. (continued)

Gene[a]	NRF1 site	Human chromosome	Circadian regulatory function	Reference
RORA	none	15q22.2	A retinoid-related orphan receptor; activates BMAL1 transcriptional activity through RORA binding box	Akashi & Takumi, 2005
RORC	none	1q21	Also a retinoid-related orphan receptor; mostly expressed in skeletal muscle	Hirose *et al*., 1994
CSNK1D	none	17q25	A member of the casein kinase I (CKI) gene family; maintains 24 hr circadian cycle length by phosphorylation regulation of PER proteins	Etchegaray *et al*., 2009
AANAT	none	17q25	An acetyltransferase; a penultimate enzyme in melatonin synthesis, regulated by cAMP-dependent phosphorylation	Hintermann *et al*., 1996; Coon *et al*., 2001
HLF	none	17q22	A member of the PAR TF family; a clock output gene	Young *et al*., 2001; Young *et al*., 2002
CCRN4L	none	4q31.1	Also known as Nocturnin in mouse; a deadenylase in circadian output pathways	Garbarino-Pico *et al*., 2007; Green *et al*., 2007
E4BP4	none	9q22	Phosphorylation regulation by *CSNK1D*; negatively regulates expression of *Per2*	Doi *et al*., 2004; Ohno *et al*., 2007
CSNK2A1	none	20p13	A subunit of casein kinase II (CK2); CK2 phosphorylates PER2 and regulates PER2 stability with CSNK1E	Tsuchiya *et al*., 2009
CSNK2A2	none	16q21	Also a subunit of casein kinase II, as for CSNK2A1	Tsuchiya *et al*., 2009
CSNK2B	none	6p21.3	Also a subunit of casein kinase II, as for CSNK2A1	Akten *et al*., 2003; Tsuchiya *et al*., 2009

[a] *BMAL1* is also known as *ARNTL*, *BMAL2* as *ARNTL2*, *DEC1* as *BHLHB2*, *DEC2* as *BHLHB3*, *PPARGC1A* as PGC-1α, *CCRN4L* as *NOCTURNIN*, *E4BP4* as *NFIL3*, *CSNK2A1* as CSK2alpha, *CSNK2A2* as CSK2alpha', and *CSNK2B* as CSK2beta.

Figure 3.2. NRF1 ChIP assays using SK-N-SH neuroblastoma cells. *SNURF-SNRPN* enhancer is a positive control. Abbreviations: Ab+: antibody positive; Ab-: antibody negative; TI: total input of DNA; H_2O: PCR negative control; bp: base pair.

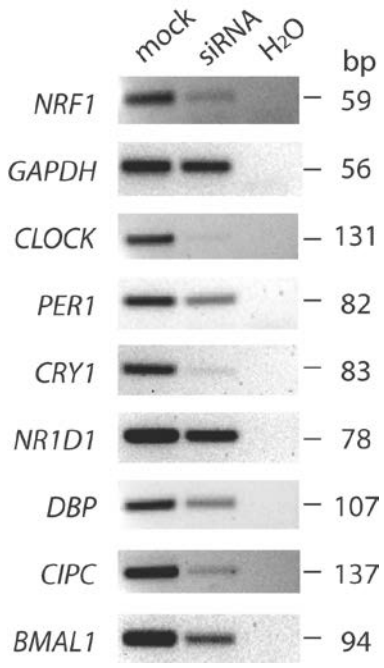


Figure 3.3. *NRF1* siRNA assays using SK-N-SH neuroblastoma cells. *GAPDH* is a negative control. Mock: no siRNA transfection negative control. siRNA: *NRF1* siRNA transfection; H_2O: PCR negative control; bp: base pair.

3.3.3 NRF1 regulates expression through the promoters of *CLOCK*, *CRY1* and *NR1D1*, and the enhancers of *PER1* and *DBP*

To test the function of NRF1 sites in the 5' regulatory regions of major circadian regulatory genes, luciferase reporter constructs for *CLOCK*, *PER1*, *CRY1*, *NR1D1* and *DBP* were made using the conserved promoter sequences of *CLOCK*, *DBP*, *CRY1*, *PER1*, and *NR1D1*, and the conserved enhancer sequences of *DBP* and *PER1* [**Figure 3.4(a), Figure 3.1**].

For the *CLOCK* promoter, two luciferase constructs were made in pGL3basic [**Figure 3.4(b)**], one with 2 intact NRF1 sites (CLOCK-pr) and the other with a site-specific mutation in the second NRF1 site [CLOCK-pr-mu, **Figure 3.4(a)**]. As expected, mutation of the second NRF1 site in the promoter significantly reduces the luciferase activity compared to the intact promoter construct in transfected SK-N-SH neuroblastoma cells [**Figure 3.4(b)**].

For the *CRY1* promoter, three constructs have been made [**Figure 3.4(c)**] with variation in the size of the promoter. Compared to the full-length *CRY1* promoter (CRY1-pr), the luciferase construct CRY1-pr-s2 with a deletion of just the NRF1 site has significantly reduced the luciferase activity (by ~30%, $P<0.01$), indicating that NRF1 is a transcription activator of the *CRY1* promoter. To examine the effects on transcriptional activity of loss of E-box motif (Ueda *et al*., 2005), as well as NF-Y and Sp1 sites, in the *CRY1* promoter, we also made a short version of *CRY1* promoter with only the NRF1 and NF-Y site [CRY1-pr-s1, **Figure 3.4(c)**]. Surprisingly, this construct showed a much higher activity than the full-length of *CRY1* promoter construct. Two potential mechanisms can explain this observation. First, either or both of two, long conserved blocks of sequence 5' to the E-box [**Figure 3.1(b)**] could bind a repressor of *CRY1* transcription, which would be removed in the CRY1-pr-s1 construct. Alternatively, the increased activity of this construct versus the full-length construct may be due to shortening the distance from the NRF1 and NF-Y sites to the TSS. Indeed, the most common position of the NRF1 motif in genome-wide bioinformatics analyses is centered at 62-nt 5' of the TSS (Xie *et al*., 2005), so the NRF1 site in the CRY1-pr-s1 construct is likely to be in a more optimal position compared to the position in the wildtype *CRY1* promoter construct.

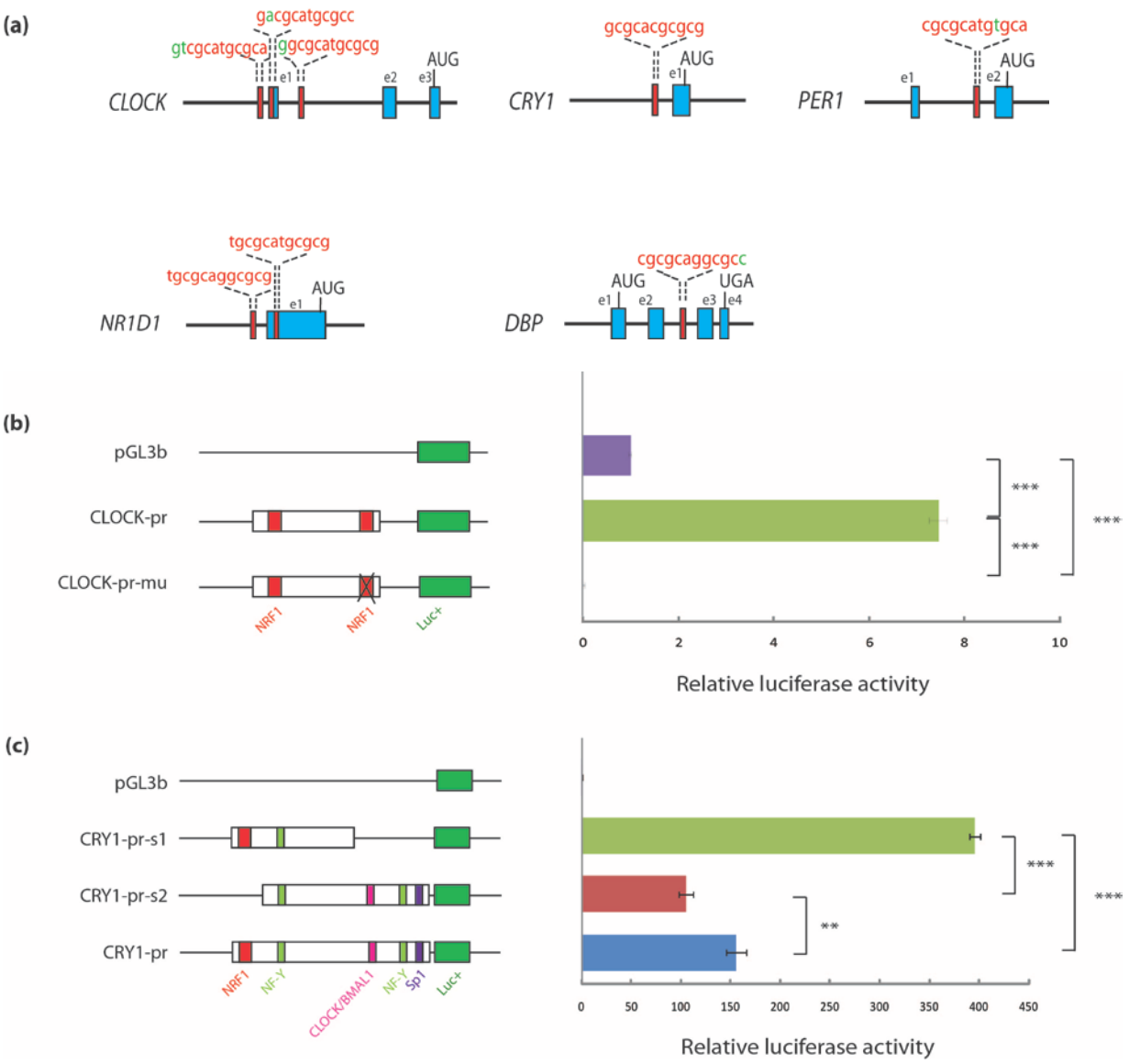


Figure 3.4. (continued below)

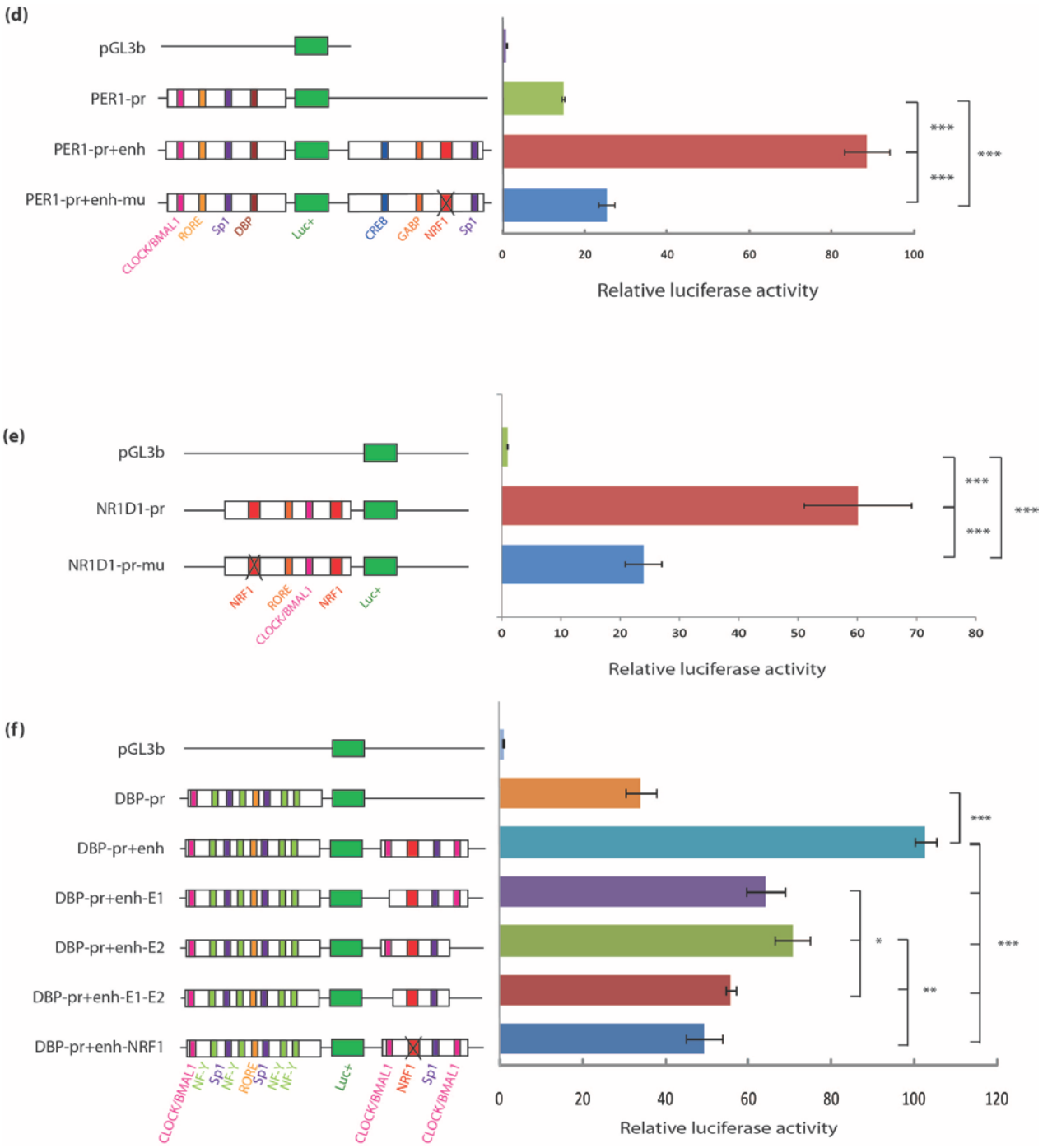


Figure 3.4. Luciferase reporter assays of 5' regulatory elements of *CLOCK*, *CRY1*, *PER1*, *NR1D1* and *DBP*. (a) Gene structure for *CLOCK*, *CRY1*, *PER1*, *NR1D1* and *DBP*, with sequence and location of NRF1 putative binding sites. For NRF1 sites, red represents nucleotides matching

to the NRF1 consensus recognition site and green represents less optimal nucleotides. Luciferase reporter assays for (b) the *CLOCK* gene promoter in SK-N-SH neuroblastoma cells, (c) the *CRY1* gene promoter in SK-N-SH neuroblastoma cells, (d) the *PER1* gene 5' regulatory element in SK-N-SH neuroblastoma cells, (e) the *NR1D1* gene promoter in NIH3T3 cells and (f) the *DBP* gene 5' regulatory element in NIH3T3 cells. $P<0.05$: *; $P<0.01$: **; $P<0.001$: ***.

Similarly, *PER1* 5' regulatory elements were cloned into the luciferase vectors [**Figure 3.4(d)**]. Compared to the *PER1* promoter construct (PER1-pr), the luciferase activity of a construct (PER1-pr+enh) carring the *PER1* promoter with an intact putative intron 1 enhancer increased by approximately 5.9-fold [$P<0.001$, **Figure 3.4(d)**] in transfected SK-N-SH neuroblastoma cells. This result confirms our hypothesis that the *PER1* intron 1 NRF1 site is within an enhancer element for *PER1* promoter activity. Not surprisingly, a single site-directed mutation of the NRF1 binding site in the *PER1* enhancer (PER1-pr+enh-mu construct) significantly reduced luciferase activity by approximately 70%, compared to the intact promoter and enhancer construct [$P<0.001$, **Figure 3.4(d)**].

For the *NR1D1* promoter, two luciferase constructs were made [**Figure 3.4(e)**], one with two intact NRF1 sites (NR1D1-pr) and the other with a site-specific mutation of the first NRF1 site (NR1D1-mu). As expected, the construct with the NRF1 site mutation has significantly reduced the luciferase activity by approximately 60% compared to the intact promoter construct in transfected NIH3T3 cells [$P<0.001$, **Figure 3.4(e)**].

Previously, two E-box elements in intron 2 of *DBP* were shown to be critical for circadian transcriptional regulation of this gene in the SCN by CLOCK/BMAL1 (Ripperger et al., 2000; Ripperger and Schibler, 2006). Deletion of the two E-box elements and sequence in between in Rat-1 cell line and in a mouse model showed loss of circadian mRNA expression of *DBP* compared to the wildtype and altered histone modification at the chromatin level. By bioinformatics analyses, we found a conserved, putative NRF1 site between the two E-box elements in intron 2 of *DBP* [**Figure 3.4(a), (f), Figure 3.1(g)**]. To test the role of the NRF1 site in this region, we cloned the *DBP* promoter alone (DBP-pr construct) and with the enhancer element (DBP-pr+enh) into a pGL3basic luciferase reporter vector. As expeced, the *DBP* promoter construct is functional and has increased the luciferase activity by 34-fold ($P<0.01$) compared to the pGL3basic vector, while the enhancer construct (DBP-pr+enh) has increased the transcriptional activity another 3-fold to the *DBP* promoter construct [overall 102-fold compared to the pGL3basic, $P<0.001$, **Figure 3.4(f)**]. Deletion of each individual (DBP-pr+enh-E1, DBP-pr+enh-E2) or both E-boxes (DBP-pr+enh-E1-E2) within the enhancer significantly reduced luciferase and hence enhancer activity by 38%, 32%, and 47%, respectively [$P<0.001$, $P<0.001$, $P<0.001$, respectively, **Figure 3.4(f)**]. Interestingly, mutation of the NRF1 binding site with both E-boxes intact (DBP-pr+enh-NRF1) reduced *DBP* enhancer activity to a greater extent [by 52%, $P<0.0001$, **Figure 3.4(f)**] than mutation of either E-box motif and at least as much as mutations of both E-boxes [**Figure 3.4(f)**]. These data provide strong evidence that NRF1 acts as an activator of the *DBP* gene enhancer.

As a complementary approach, we used the intact *DBP* promoter+enhancer luciferase construct (DBP-pr+enh) and co-transfected with expression vectors for NRF1 (NRF1-VP16; Ramachandran *et al*., 2008), CLOCK and BMAL1 (Kondratov *et al*., 2006). CLOCK and

BMAL1 increased luciferase activity of the DBP-pr+enh construct by ~18-fold (P<0.001, **Figure 3.5**). Similarly, NRF1 alone or in combination with CLOCK and BMAL1 increased luciferase activity of the DBP-pr+enh construct by ~14 and ~21 fold, respectively (P<0.001, **Figure 3.5**). These data demonstrate that NRF1 as well as the CLOCK-BMAL1 heterodimer activate the *DBP* intron 2 enhancer.

Taken together, luciferase reporter assays have demonstrated that NRF1 is a major transcriptional activator for *CLOCK*, *CRY1*, and *NR1D1* though promoter target sites. Similarly, NRF1 acts as a major regulator of *PER1* and *DBP* transcription through intronic enhancer sites.

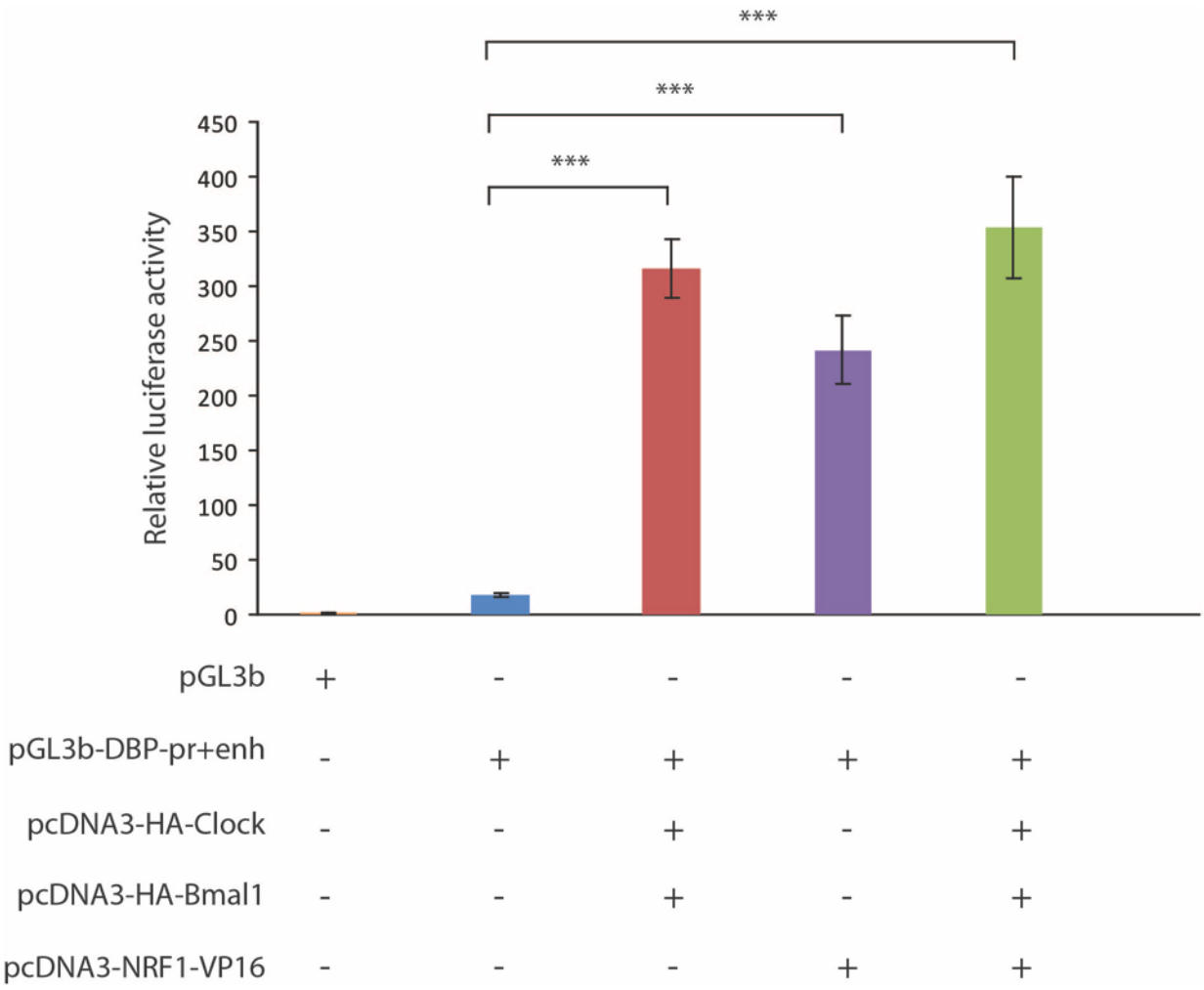


Figure 3.5. Co-transfection of the *DBP* promoter+enhancer luciferase construct with expression vectors for CLOCK, BMAL1 and NRF1. *P*<0.001: ***.

3.3.4 NRF1 and the CLOCK/BMAL1 heterodimer are in a molecular complex

We have shown that NRF1 binds spatially close to two CLOCK/BMAL1 binding sites in the *DBP* intron 2 enhancer and together NRF1 and CLOCK/BMAL1 co-regulate this *DBP* circadian enhancer. Therefore, we hypothesize that NRF1 interacts with CLOCK/BMAL1 in a molecular complex, and that CLOCK may acetylate NRF1 with deacetylation by SIRT1. In support of this hypothesis are the following observations: 1) CLOCK is known to acetylate histone H3, histone H4, and BMAL1 (Doi *et al*, 2006; Hirayama *et al*, 2007); 2) SIRT1 is known to associate with CLOCK/BMAL1 and to promote the deacetylation of histone H3, BMAL1, and PER2 (Asher *et al*, 2008; Nakahata *et al*, 2008) and 3) *in vitro* acetylation of NRF1 by pCAF (p300/CBP-associated factor) enhances its function (Izumi *et al*., 2003), suggesting that NRF1 acetylation *in vivo* may contribute to its mechanism of action at least at a subset of *cis*-regulatory elements that it targets.

The acetylated lysine residue(s) in NRF1 [**Figure 3.6(a)**] are not known (Izumi *et al*, 2003), and other HAT enzymes, including CLOCK, might be relevant for NRF1 acetylation *in vivo*. Few targets of the CLOCK [**Figure 3.6(b)**] HAT enzymatic function have been identified (Doi *et al*., 2006; Hirayama *et al*., 2007). For histone H3, it is lysine 14 (K14) that is acetylated [Doi *et al*, 2006, **Figure 3.6(c)**], while phosphorylation of histone H3 at S10 (which is induced by light, Crosio *et al*, 2000) enhances acetylation at K14 (Doi *et al*., 2006). For BMAL1, it is K537 (mouse) that is acetylated [**Figure 3.6(c)**], whereas the acetylated lysines in histone H4 have not been identified (Doi *et al*., 2006). Our comparison of histone H3 and BMAL1 amino acid sequences around the CLOCK-acetylated lysine identified a shared SXGGK motif with similar motifs present in a region of more extended similarity between histone H4 and NRF1 [SXGRGK;

Figure 3.6(c)]. This observation predicts lysine 221 in NRF1 as a candidate for potential CLOCK acetylation.

To test this hypothesis, co-immunoprecipitation was performed to examine a potential interaction between NRF1 and CLOCK. We used antibodies against NRF1 and, as a positive control, antibodies against BMAL1 for immunoprecipitation and then detected CLOCK proteins on a western blot [**Figure 3.6(d)**]. As expected (Lee *et al*., 2001; Tamaru *et al*., 2003; Kondratov *et al.*, 2006), BMAL1 co-immunoprecipitated both native and phosphorylated CLOCK-isoforms from nuclear extracts. Intriguingly, NRF1 not only co-immunoprecipitated both CLOCK isoforms, but preferentially associated with the phosphorylated, active form of CLOCK [**Figure 3.6(d)**]. Besides full length CLOCK and phosphorylated CLOCK isoforms, we detected a low level of a smaller CLOCK protein in nuclear extract, in NRF1-IP and BMAL1-IP samples [**Figure 3.6(d)**]. According to previous studies, a smaller *Clock* mRNA isoform was found in mouse with alternative splicing of exon 18, which leads to an in-frame 30 amino acid deletion (King *et al*., 1997). Using primers located in exons 15 and 21 (**Table 3.4**), RT-PCR was performed using cDNA from NIH3T3 cells (see below). As expected, we found two RT-PCR bands [**Figure 3.6(e)**], the sizes which match the more abundant full-length and the less abundant alternatively spliced (lacking exon 18) *Clock* mRNA isoforms. Therefore, all three protein bands detected by western analysis are different isoforms of CLOCK.

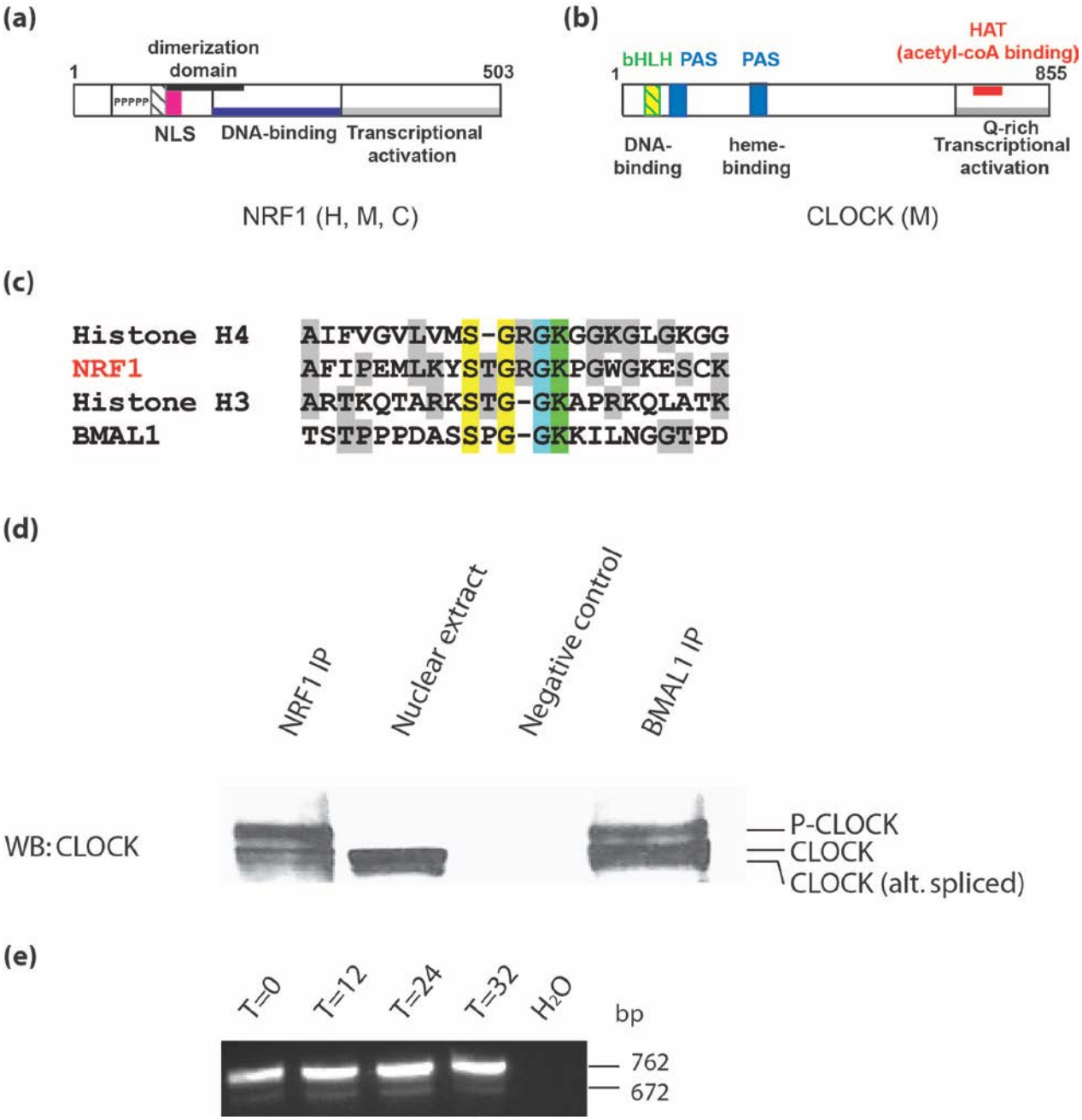


Figure 3.6. Co-immunoprecipition of NRF1 or BMAL1 and CLOCK, as detected by western blot. (a) Structure of the conserved NRF1 polypeptide. NRF1 from human (H), mouse (M), and chicken (C) are 99-100% identical over the entire sequence. The position of structural domains is indicated, including nuclear localization signal (NLS, pink vertical box), DNA binding (dark blue

horizontal box), dimerization (black horizontal bar) an transcriptional activation (grey box) domains. NRF1 has 5 serine residues that are phosphorylated (P). Modified from Scarpulla (2008). **(b)** Structure of the CLOCK polypeptide from mouse (M) showing functional motifs. **(c)** Amino acid alignment of CLOCK-targeted acetylation motifs in histone H3 and BMAL1, and similar motifs in histone H4 and NRF1. The acetylated lysine in H3 and BMAL1 have green shading, with residues present in all 4 sequences in green, blue or yellow shading, while gray shading denotes amino acids present in 2 or 3 of the sequences. **(d)** Detection of CLOCK protein on a western blot, showing native CLOCK, phosphorylated CLOCK (P-CLOCK), and a low level of CLOCK resulting from alternatively spliced (alt. spliced) *Clock* mRNA (see **e**). Abbreviations: IP: immunoprecipitation; nuclear extract: total protein extract from nucleus; negative control: the negative control in immunoprecipitation uses control agarose resin. **(e)** RT-PCR analysis of full-length (762-bp) and alternatively-spliced (672-bp) *Clock* mRNA isoforms. Each lane shows RT-PCR products from NIH3T3 cell cDNA derived from different time points after serum-shock.

3.3.5 Ultradian oscillations of *Nrf1* mRNA and NRF1 protein in serum-shocked NIH3T3 cells

It has been reported that a subset of clock genes can be synchronized and subsequently undergo at least two circadian oscillation cycles in cell line models (Balsalobre *et al*., 1998; 2000; Earnest & Cassone, 2005). To test if *Nrf1* mRNA levels oscillate in a circadian or other fashion, we used serum-shocked NIH3T3 cells, an established model to synchronize circadian cycles of a subset of clock genes (Osland *et al*., 2010). As expected, both *Dbp* and *Nr1d1* showed typical

rhythmic expression in this cell model, with the peak phase at 24 hr and 20 hr, respectively, after serum-shock [**Figure 3.7(a)&(b),** Balsalobre *et al*., 1998], while *Clock* and *Per1* mRNA levels did not cycle (data not shown). These findings are consistent with previous reports with respect to peaks and phases of oscillation and oscillatory behavior of mRNAs encoding known circadian regulators (Balsalobre *et al*., 1998; Osland *et al*., 2010). Interestingly, *Nrf1* mRNA levels oscillated [**Figure 3.7(c)**], with the peak phase at 16 hr and 32 hr, respectively, after serum-shock, indicating an ultradian rhythm (~16 hr; Hughes *et al*., 2009). At the protein level, NRF1 showed a similar ultradian rhythm as for mRNA levels, when compared to the positive control (DBP) and the β-actin negative control [**Figure 3.7(d) & (e)**]. As we have shown that NRF1 controls expression of circadian regulatory genes in both the positive and negative feedback loops, it is not surprising that NRF1 does not cycle in a typical circadian manner. Indeed, even at the nadir of *Nrf1* mRNA and NRF1 protein levels, there are still significant levels present in these cells under these conditions [**Figure 3.7(c),(e)**]. Further studies are needed to determine if NRF1 levels oscillate *in vivo* in the SCN and other tissues.

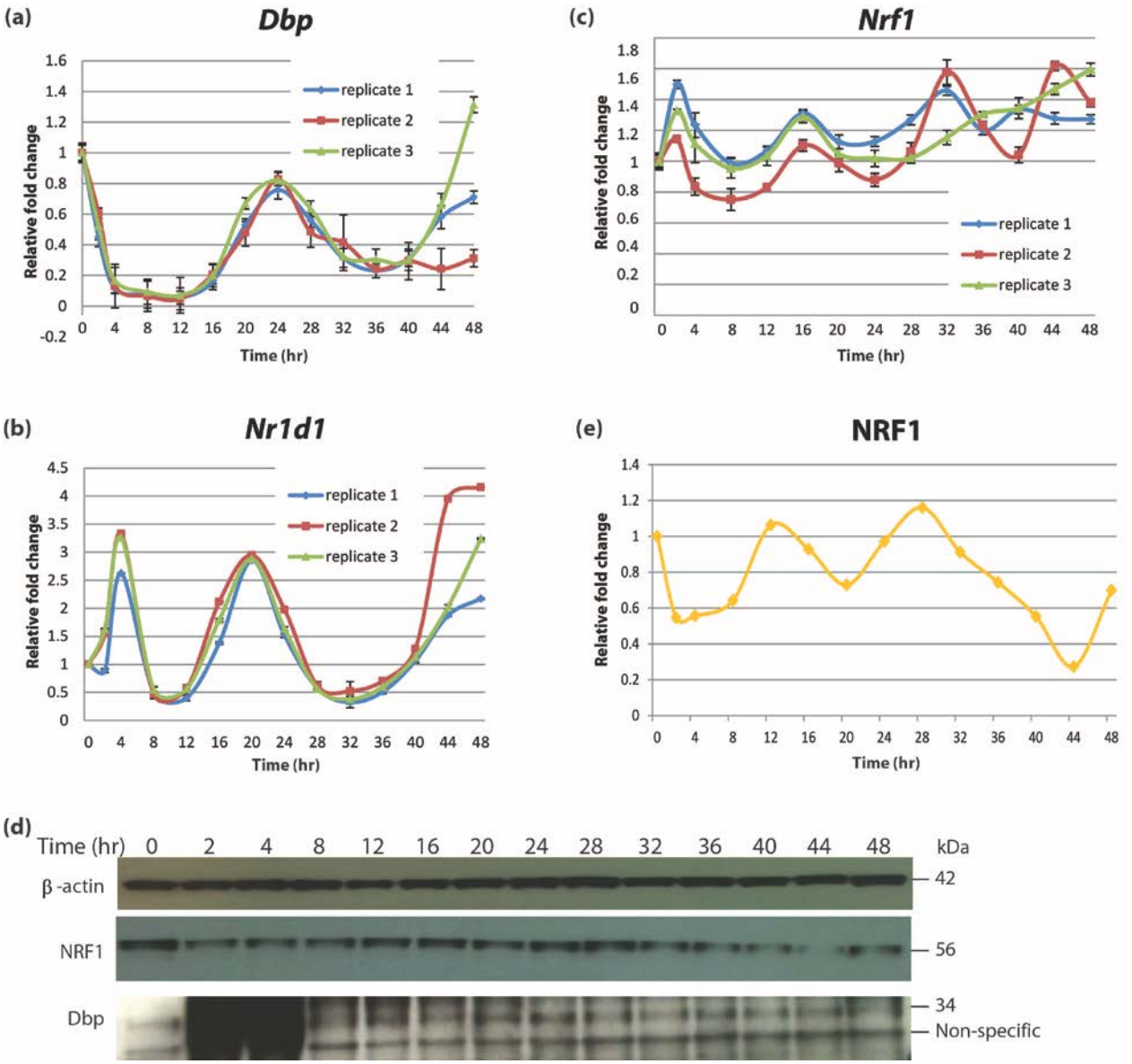


Figure 3.7. NRF1 mRNA and protein expression levels in serum-shocked NIH3T3 cells. (a) mRNA expression levels of *Dbp*. **(b)** mRNA expression levels of *Nr1d1*. **(c)** mRNA expression levels of *Nrf1*. Three different biological replicates are shown. **(d)** Protein expression of DBP and NRF1 detected on western blot. β-actin was used as a control. **(e)** Quantification of protein amounts in each lane for NRF1 normalized to β-actin, using Image J software.

3.4 DISCUSSION

In this study, for the first time NRF1 has been demonstrated to positively regulate numerous circadian genes, including core regulatory genes and CCGs with regulatory roles (**Figure 3.2-3.4, 3.8, Table 3.3**). Overall, ~56% of these genes are targets of NRF1 regulation. Furthermore, NRF1 and the active, phosphorylated CLOCK polypeptide interact in a molecular complex. These observations indicate a significant impact of NRF1 on circadian biology. In particular, we used numerous molecular genetic methods to show that NRF1 positively regulates transcriptional activity of *CLOCK*, *CRY1*, and *NR1D1* through promoter sites, as well as *PER1* and *DBP* through enhancer sites (**Figure 3.8**). Intriguingly, we also observed that *BMAL1*, which was not predicted to be a direct target of NRF1, had down-regulated mRNA levels in *NRF1* siRNA assays. This could be caused by down-regulation of other NRF1 target genes in the circadian system that encode regulators of *BMAL1* expression (**Table 3.3**, **Figure 3.8**) or an unknown circadian factor that is also under NRF1 regulation and remains to be identified. If additional circadian regulatory genes that are not direct targets if NRF1 show a similar secondary effect, then the role of NRF1 in the circadian network would be extremely broad.

NRF1 was first identified with a major role in activating nuclear genes encoding mitochondrial functions and biogenesis (Virbasius *et al*., 1993; Scarpulla, 2008), in cell cycle regulation (Cam *et al*., 2004; Patti *et al*., 2003), neuronal development and function (Chang *et al*., 2005; Yang *et al*., 2006; Liang& Wong-Riley, 2006, Smith *et al*., 2004), in Prader-Willi syndrome (Stefan M, *et al*., in preparation; **Chapter 2**), and is implicated in diabetes (Patti *et al*., 2003; Mootha *et al*., 2003; Gaulton *et al*., 2008; Liu *et al*., 2008; He *et al*., 2008). Studies of NRF1 orthologs have shown that loss of Nrf1 function in zebrafish is embryonic lethal and leads to a specific loss of photoreceptor neurons and their precursors (Becker *et al*., 1998). As circadian

rhythms are entrained by light through retinal afferents to the SCN (Rutter *et al*., 2002), NRF1 could regulate the process of light entrainment and neuronal connections to the SCN, although this hypothesis remains to be examined. Additionally, a serum shock induces circadian gene expression in cell lines (Balsalobre *et al*., 1998), and NRF1 is in an inactive state in serum-starved cells but becomes active (phosphorylated) with serum addition (Gugneja & Scarpulla, 1997). Finally, in the peripheral system, PPARGC1A, a transcriptional coactivator to NRF1 and other TFs in mitochondrial metabolism, is rhythmically expressed in the mouse liver and skeletal muscle, and may play a role in metabolic adaptations to the daily circadian cycle (Liu *et al*., 2007). These observations are consistent with our findings that NRF1 has major regulatory roles in the circadian system.

The molecular mechanism of circadian transcriptional regulation is best understood for the *DBP* gene, highly expressed and strongly rhythmic in the SCN and peripheral tissues such as the liver (Lopez-Molina *et al*., 1997; Yan *et al*., 2000), which involves the rhythmic binding of CLOCK/BMAL1 heterodimers to E-box elements in an intron 2 enhancer in coordination with rhythmic changes in histone modification (Ripperger *et al*., 2000; Ripperger & Schibler 2006; Kiyohara *et al*., 2008). Additionally, CRY1 can also be recruited to the non-canonical E-box in the *DBP* and cause delay phrase of *DBP* expression in liver (Stratmann *et al*., 2010). In this study, we further dissected the *cis*-acting *DBP* enhancer and identified a significant involvement of NRF1, in which the NRF1 site closely flanks the two E-boxes in intron 2 of *DBP* [**Figure 3.1(g)**]. Indeed, NRF1 affects the amplitude of transcriptional activity due to *DBP* enhancer activity to the same degree as the effect triggered by CLOCK/BMAL1 heterodimers in unstimulated NIH3T3 cells [**Figure 3.4(f) & Figure 3.5**]. Although in transfection experiments the combination of NRF1, CLOCK and BMAL1 was not dramatically additive, it is possible that the *DBP*

transcriptional capacity in the luciferase reporter system is saturated by NRF1 alone or CLOCK-BMAL1 alone. Alternatively, NRF1 and CLOCK/BMAL1 may have independent roles in activation of *DBP* enhancer function. The latter seems unlikely given that we have identified NRF1 and the active CLOCK transcription factor in a molecular complex, making it likely that NRF1 and CLOCK function together in a complementary manner. Interestingly, the amplitude of transcriptional activity (Dibner *et al*., 2009; O'Neill, 2009) plays a role in the period of the circadian clock, so NRF1 may contribute to this aspect of circadian function by its ability to strongly activate transcription through enhancer and promoter mechanisms.

By protein immunoprecipitation studies we have shown that NRF1 and CLOCK/BMAL1 interact in a molecular complex, and it is striking that NRF1 preferentially interacts with the phosphorylated, active isoform of CLOCK in the nucleus [**Figure 5(c)**; Lee *et al*., 2001; Tamaru *et al*., 2003; Shim *et al*., 2007]. It is known that the status of CLOCK phosphorylation is dynamic, with phosphorylation of CLOCK associated with its translocation from the cytoplasm to the nucleus (Kondratov *et al*., 2003; Shim *et al*., 2007) and an enhancement of its E-box binding ability, whereas hyperphosphorylation decreases protein stability of CLOCK (Yoshitane *et al*., 2009; Spengler *et al*., 2009). Furthermore, many factors can transiently interact with CLOCK/BMAL1 for specific functions, for example, CIPC-mediated CLOCK phosphorylation (Yoshitane *et al*., 2009), SIRT1-mediated deacetylation of BMAL1 and PER2 (Asher *et al*., 2008; Nakahata *et al*., 2008; Masri & Sassone-Corsi, 2010), and MLL1 (mixed lineage leukemia-1)-mediated methylation of H3K4 (Masri & Sassone-Corsi, 2010). Therefore, it is plausible that the interaction between NRF1 and CLOCK/BMAL1 also occurs in a transient manner at the promoter or enhancer elements of a subset of genes involved in circadian rhythms.

Interestingly, our data showed that peak levels of NRF1 mRNA and protein oscillate with an ~16 hr ultradian rhythm (~1.5th harmonic cycle, Hughes *et al*., 2009) in serum-shocked NIH3T3 cells [**Figure 3.7(c)**]. It is not surprising that NRF1 levels oscillate, as in other work (see **Chapter 4**), we found that NRF1 is autoregulated by a feed-forward mechanism, although the mechanism of negative feedback regulation recquisite for an oscillatory system is not known. One candidate would be a repressive transcriptional mechanism (see **Chapter 4**). Another mechanism that could negatively regulate *NRF1* mRNA and protein levels would be miRNA targeting, and our recent studies have found that miR-182 and miR-96 are involved in down-regulation of NRF1 (B. J. Henson, R. D. Nicholls, unpublished data). This is intriguing as the miR-182/miR-96 family of miRNAs has been suggested to show circadian expression in retinal neurons (Xu *et al*., 2007) although more recent studies suggest that the oscillatory activation of miR-182/miR-96 in retinal photoreceptor neurons is due to activation by light with rapid turnover in the dark (Krol *et al*., 2010). These and additional mechanisms acting in retinal (Becker *et al*., 1998) and/or SCN neurons could thus play a role in regulation of NRF1 and subsequent circadian regulation of downstream NRF1 target genes. Combined with our data showing that NRF1 targets include genes from both positive and negative feedback loops (**Table 3.3**), we propose that the circadian system can interpret the oscillating levels of NRF1 and its interaction with the CLOCK/BMAL1 complex to regulate the affinity and binding at *cis*-regulatory elements of target genes to integrate the regulatory networks of both the positive and negative circadian feedback loops.

Genome-wide microarray analyses have shown that 10-15% of genes are CCGs with circadian expression and that many of these genes are involved in metabolism and energy balance (Reppert & Weaver, 2002; Mendoza, 2007; Masri & Sassone-Corsi, 2010). Interestingly, energy balance in the body is sensed and regulated by the AMP-activated protein kinase (AMPK, Hardie,

2008; Spasic *et al*., 2009; Richter *et al*., 2009; Lim *et al*., 2010), which regulates the circadian clock through phosphorylation and stability of CRY1 and entrains circadian function to metabolism (Lamia *et al*., 2009; Suter & Schibler, 2009). AMPK also activates NRF1-pathways (Bergeron *et al*., 2001; Garcia-Roves *et al*., 2008), and provides a link between energy balance, diabetes (Hardie, 2008), exercise (Ojuka *et al*., 2004), and NRF1 levels and activity (Timmons *et al*., 2006; Wright *et al*., 2007). Additionally, PPARGC1A is one of several co-activators for NRF1, particularly for nuclear genes encoding mitochondrial functions (Scarpulla, 2008), and PPARGC1A is known to play a key role in integration of circadian function (Liu *et al*., 2007) and metabolism (Cunningham *et al*., 2007). Collectively, this provides a link of NRF1 to energy balance and metabolism, each of which can critically influence circadian rhythms.

In vertebrates, light entrains the developing circadian clock via photoreceptive retinal ganglion cells that express melanopsin and that project to the SCN (Vallone *et al*., 2007). Intriguingly, mutation of *NRF1* ortholog in zebrafish demonstrated an essential role in development of all types of photoreceptors (Becker *et al*., 1998). The *NRF1* ortholog in fly (*EWG*) is also essential for neuronal development (DeSimone & White, 1993). Supporting evidence from the Allen brain atlas (Jones *et al*., 2009) by *in situ* hybridization shows enriched expression of *Nrf1* in the SCN of the hypothalamus in mouse, as well as hypothalamic nuclei involved in the regulation of feeding, and other non-hypothalamic neuronal structures. This observation, combined with our data showing NRF1 contributes to enhancer function of the SCN-specific *DBP* enhancer (Ripperger et al., 2000), as well as other key genes for SCN circadian function such as *CLOCK*, *PER1*, *CRY1*, *NR1D1*, and others, supports a hypothesis for a role of NRF1 in gene regulation of the central pacemaker of circadian rhythms. Future studies will determine the roles

and mechanisms of action of NRF1 in establishment and/or function of neural and molecular pathways underlying circadian rhythms.

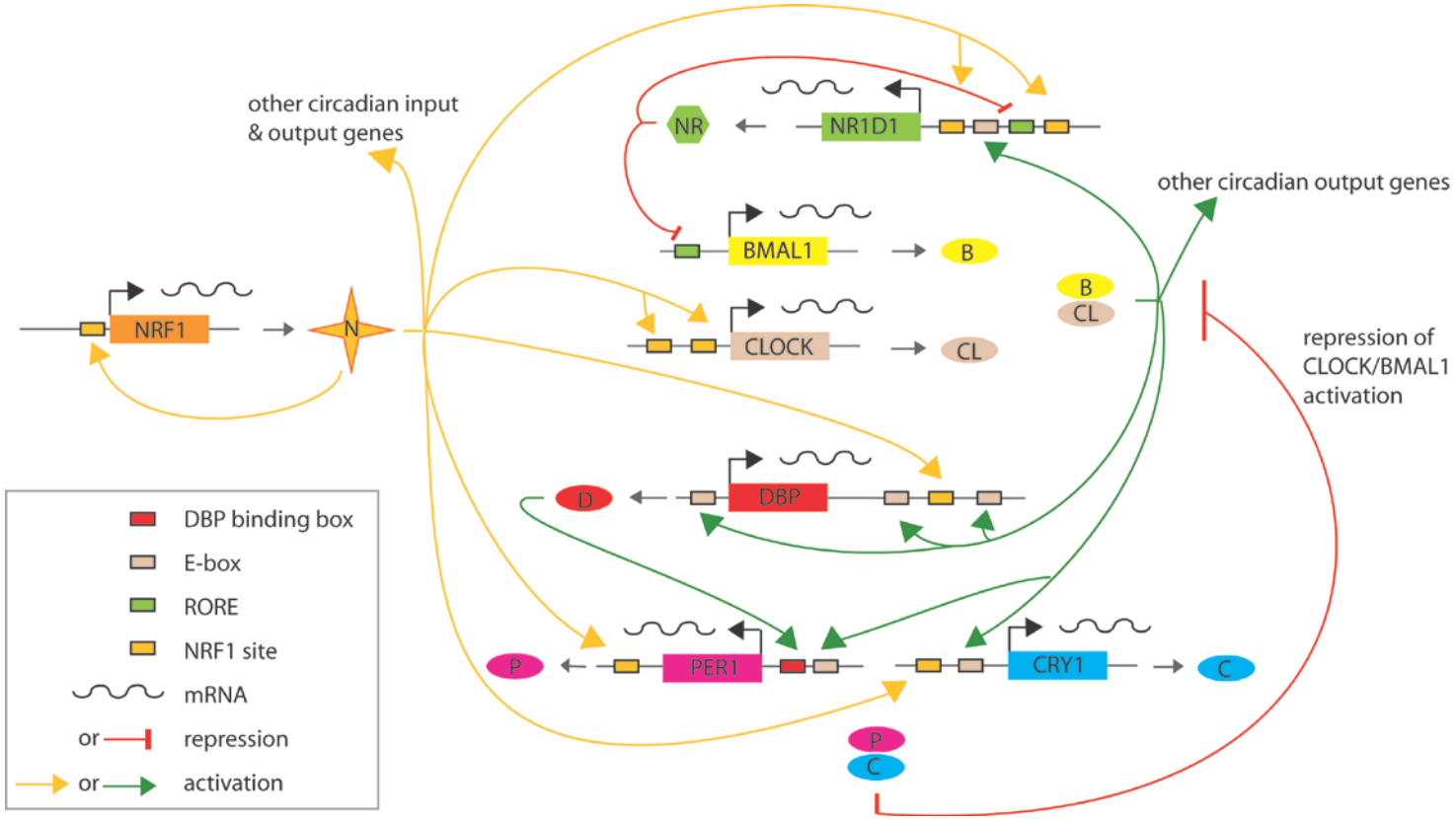


Figure 3.8. Summary of NRF1 regulation in the circadian system. The model depicts the canonical clock mechanism involving positive and negative feedback loops comprised of CLOCK/BMAL1 heterodimers binding to E-box elements in target genes and subsequent repression of these pathways by PER/CRY heterodimers. NR1D1 provides an ancillary feedback loop, while DBP encodes an output regulatory TF and also regulates PER1. Post-translational regulation by phosphorylation and protein degradation are not shown for simplicity. NRF1 regulation occurs for many components of the circadian regulatory hierarchy.

4.0 ADDITIONAL MECHANISMS AND SYSTEMS OF NRF1 REGULATION

4.1 INTRODUCTION

4.1.1 Hereditary spastic paraplegia

Hereditary spastic paraplegias (HSP) are neurodegenerative disorders characterized primarily by progressive spasticity in the lower limbs due to axonal degeneration of the corticospinal tract (CST). Other neural-related clinical features include progressive symmetric spasticity of lower extremities, often but not always with muscle weakness, and increased stiffness, hyperreflexia, extensor plantar responses and, in some cases, bladder disturbances and vibratory sense impairment (Depienne *et al*., 2007; Fink 2009). Dominant inheritance accounts for ~70% of HSP and mutations in one gene, *SPAST* (*SPG4*), occur in about 40% of these cases. However, to date, 45 autosomal dominant, recessive and X-linked gene loci are chromosomally linked to HSP pathogenesis, with etiological gene mutations identified for only 22 of these (Depienne *et al*., 2007; Salinas *et al*., 2008; Dion *et al*., 2009; Fink, 2009). Selection of new candidate genes that affect function and trafficking in long axons is difficult as known hereditary spastic paraplegia (HSP) genes encode proteins involved in a multitude of biochemical pathways including known or potential roles in mitochondria, microtubules, anterograde transport, endosomal, endoplasmic reticulum (ER) and Golgi, as well as additional functions such as cytoplasmic, gap junction, and myelin sheath or Schwann cell function (Salinas *et al*., 2008; Dion *et al*., 2009; Park *et al*., 2010).

Our knowledge of *cis*-binding motifs and role of NRF1 in transcriptional regulation in the PWS region (**Chapter 2**) provided a basis for initially examining *SPG6* (*NIPA1*, on human chromosome 15q11.2), followed by the *Spastin* (*SPG4*) gene promoter since *SPG4* is most frequently involved in spastic paraplegia. Having identified potential NRF1 binding sites in the promoters of these two genes (see Results), we extended our analyses to all 22 cloned HSP loci. We hypothesize that HSP loci are co-regulated by a common GRN in order to determine proper spatiotemporal expression patterns in the motor cortex to accomplish specific CST functions.

4.1.2 NRF1 autoregulation

NRF1 is known to be crucial for the expression of nuclear genes that encode mitochondrial products (Scarpulla, 2008), genes in cell cycle regulation (Morrish *et al*., 2003; Cam *et al*., 2004), development (Chen *et al*., 1997; Efiok & Safer, 2000; Moriuchi *et al*., 1997; Solecki *et al*., 2000) and neural functions (Chang & Huang, 2004; Smith *et al*., 2004; Chang *et al*., 2005; Yang *et al*., 2006). Indeed, loss of Nrf1 function is lethal across the animal kingdom (**Chapter 1.1.1**). Previous reports have shown that some TFs are autoregulated by binding to their own promoters. For example, NR1D1 has a Rev-erbα response element (RORE) on its own promoter [**Figure 3.1(e)**], which has been shown to mediate transcriptional repression when NR1D1 protein level goes up in an *in vitro* system, implying this gene is potentially autoregulated (Adelmant *et al*., 1996). YY1 has a cluster of YY1 binding sites on its intron 1 as a transcriptional activator (Kim *et al*., 2009). When exogenous YY1 was overexpressed in an induced cell line, endogenous YY1 protein level went down to maintain homeostasis through regulation by the cluster of YY1 sites intron 1, proving the mechanism of YY1 autoregulation (Kim *et al*., 2009). These data suggested a TF autoregulatory system is functioned by a transcriptional feedback loop

to repress its own expression. Intriguingly, we found two highly conserved NRF1 sites in the 5' regulatory region and one site in intron 1 of the *NRF1* gene (**Figure 4.1**). All putative NRF1 sites are highly conserved from fish to human (**Figure 4.1**). Therefore, we hypothesize that NRF1 autoregulates its own transcriptional activity.

4.1.3 NRF1 binding site studies

As mentioned in Chapter **1.1.4**, several genome-wide studies by bioinformatics analysis or experiments (ChIP-chip) have revealed that approximately 5-6% of the genes are regulated by NRF1. Functional NRF1 binding sites have been confirmed experimentally with sequence variance of 1 nucleotide in either GCGC and/or 1 nucleotide in the AY segment (Stefan *et al*., in preparation, **Figure 2.2 & 3.1**). These studies also suggested that multiple NRF1 binding sites may also increase binding affinity by cooperative binding even if individual recognition motifs may not be optimal.

To further study the parameters that affect NRF1 binding ability, an artificial model has been designed. In this model, six minigenes have been synthesized in pIDTSMART vectors (IDT), each containing 10, 8, 6, 4, 2, or 1 copies of NRF1 consensus recognition sites. These minigenes were subcloned from the pIDTSMART plasmids into the promoter position of a modified pGL3basic luciferase reporter vector. Subsequently, these series of pGL3basic-derived vectors were transfected into SK-N-SH neuroblastoma cells to compare the luciferase activities. These studies will not only improve bioinformatics methods for identification of functional NRF1 sites genome-wide, in any animal species, but will also provide critical knowledge and reagents to aid structural analyses of NRF1 as well as the generation of novel reagents for future *in vivo* functional studies.

Figure 4.1. Multi-sequence alignment of the *NRF1* promoter-exon 1-intron 1 region. NRF1 sites are highlighted in bold red. A YY1 site is highlighted in bold pink. Sp1 sites are highlighted in bold purple. The exon/intron boundary is highlighted in bold blue.

4.2 MATERIALS AND METHODS

4.2.1 Bioinformatics and phylogenetic analyses

NRF1 sites were predicted using bioinformatics by matches on either strand to the consensus binding motif 5'-yGCGCANGCGCr-3', allowing one mismatch in one but not both GCGC motifs (y = t or c; r = a or g; neither of these two positions is required but generates a more optimal motif). Sites with one mismatch in one GCGC motif are defined as an NRF1-like site (L); various substitutions lead to differing affinity for NRF1 as determined experimentally (e.g., ACGC remains high affinity with strong binding of NRF1, but GCCC has little or no affinity, although multiple NRF1 or NRF1-like sites in a promoter can increase the binding extent at even weaker sites).

4.2.2 Gene expression studies

RNA was extracted from human SK-N-SH neuroblastoma cells with Trizol (Invitrogen) according to the manufactuer's protocol. DNase I treatment was performed to remove DNA contaminants from each RNA sample. Based on RNA concentration, determined using a spectrophotometer, 1 µg of RNA from each sample was transcribed into cDNA by SuperScript® III reverse transcriptase according to the manufacturer's protocol (Invitrogen). Regular PCR was performed to test gene expression of a subset of HSP genes using primers from **Table 4.1**.

Table 4.1. Human QPCR primers for HSP genes.

Target gene (primer)[a]	Primer number	Exon	Primer sequences
SPG3A F	RN3632	7-8	5'-TCAAGGTCTCAGGGAACCAGC-3'
SPG3A R	RN3633	8	5'-CCAGGATGAGGTAGCAGAAAAC-3'
SPG4 F	RN3430	15	5'-GACTGATGGATACTCAGGAAG-3'
SPG4 R	RN3431	16	5'-GGCAGACATATTCTTCACCTG-3'
SPG5 F	RN3613	4	5'-TTTCCCATCCACCTCACCAG-3'
SPG5 R	RN3614	5	5'-CCCGGTCTCTGAACTGAGAG-3'
SPG6 F	RN3024	4	5'-GCTCCGTCGTGCTGATTAT-3'
SPG6 R	RN3025	5	5'-AGGTAGCCCACAAACACTGG-3'
SPG7 F	RN3428	15	5'-GGTGAAGCAGTTTGGGATGG-3'
SPG7 R	RN3429	16	5'-CTTGGCCACCAGCAGTCTTG-3'
SPG8 F	RN3432	27	5'-GCCAGTTTATCTGCTCCACG-3'
SPG8 R	RN3433	28	5'-GCACCCACAACATCTGCAGG-3'
SPG10 F	RN3634	9-10	5'-CTGGCTGAGGGCACTAAAAG-3'
SPG10 R	RN3635	10	5'-TGAGCAACAGATGAACATAGTC-3'
SPG11 F	RN3624	7	5'-GCTGATGGACTGTATTCTGG-3'
SPG11 R	RN3625	8	5'-AATCTGGAACAAAATCGTAGGC-3'
SPG13 F	RN3434	10	5'-GATGCATTCCAGCCTTGGAC-3'
SPG13 R	RN3435	11	5'-CAAAGATCCTTCAACACCTGC-3'
SPG15 F	RN3636	4	5'-TTGTACTGGAGAAATGGTTGGC-3'
SPG15 R	RN3637	5	5'-TCTGACAATAAAAGAAACTCAAGC-3'
SPG17 F	RN3638	5	5'-CCTCCTGCTATTTGGCTTTGC-3'
SPG17 R	RN3639	6	5'-CTGTGGATCTCAATGATCGCTC-3'
SPG20 F	RN3438	7	5'-CATGTCAAGAAGCATGGAAGC-3'
SPG20 R	RN3439	8	5'-AGCTGCACATTCCAATCCTTG-3'
SPG21 F	RN3640	3	5'-GCTCTGACTGGATGGGGTTAC-3'
SPG21 R	RN3641	4	5'-CTGAATCCATCACAGAACTCG-3'
SPG31 F	RN3642	3	5'-GCACTTTTCACCACAGCAGAG-3'
SPG31 R	RN3643	4	5'-TTAGTTCATAATAGAATGGAAACC-3'
SPG33 F	RN3436	7	5'-CGGGTTCCTGAGCAAGAATG-3'
SPG33 R	RN3437	8	5'-TGAGAAGGTGGCCGAGCAG-3'

[a] Abbreviations: F: forward PCR primer; R: reverse PCR primer. Gene sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/Homo_sapiens/Info/Index).

4.2.3 Chromatin immunoprecipitation (ChIP)

Approximately 1×10^6 SK-N-SH neuroblastoma cells or marsupial cells were plated on 35 cm^2 plates with regular medium [medium for SK-N-SH cells: Alpha modified MEM + 10% FBS + 1% penicillin-streptomycin (P/S) + 1% 100 mM L-glutamine + 1% nonessential amino acid + 1% sodium pyruvate; medium for marsupial cells: DMEM+10% FBS + 1% penicillin-streptomycin (P/S)]. When cells reached 70-80% confluence, formaldehyde was added into the regular media to make the final concentration of 1% and the plates were incubated at 37°C for 10 minutes to cross link protein to DNA. After washing with 5 ml ice cold PBS once, another 5 ml ice cold PBS was added, cells were scraped from the plates by a cell lifter, put into a 15 ml tube, centrifuged, re-suspended in SDS lysis buffer (ChIP assay kit, Millipore), and sonicated to shear DNA. Samples were precleared with protein G-agarose/salmon sperm beads (Millipore). Protein-DNA complexes were immunoprecipitated with anti-NRF1 (from Dr. Daniel Raines, Smith *et al*., 2004) or anti-YY1 (Santa Cruz Biotechnology, sc-1703X). Complexes were collected with Protein G agarose/salmon sperm beads and washed. Protein-DNA complexes were eluted off the beads and cross-links were reversed by incubation with NaCl at 65ºC overnight. On the next day, DNA was recovered by phenol-chloroform extraction and precipitated by ethanol. PCR was performed using the recovered immunoprecipitated DNA materials as templates and primers from **Table 4.2** and **Table 4.3**. The PCR products were run on an agarose gel to examine the results.

4.2.4 *NRF1* siRNA assays

NRF1 siRNA primers were designed (Brummelkamp *et al*., 2002; Smith *et al*., 2004), cloned into the pSUPER siRNA vector and then sequenced to confirm the correct structure. SK-

N- SH cells were sub-cultured onto 75 cm^2 flasks 1 day before the nucleofection and reached 80-90% confluence. On the day of nucleofection, 4 μg of pSUPER-NRF1 DNA were transfected into 1×10^6 SK-N-SH cells using an Amaxa cuvette (Lonza) with Nucleofector® (Lonza) and the transfected cells were plated onto a 6-well plate. 24 hr after nucleofection, RNA was extracted with Trizol and treated by DNase I to remove genomic DNA contaminants. 1 μg of RNA from each sample was reverse transcribed into cDNA using SuperScript® III reverse transcriptase (Invitrogen) according to the manufacture's instruction. PCR was performed using primers for each gene (**Table 4.1**) to examine gene expression with *GAPDH* as a negative control and *NRF1* as a positive control to ensure effective down regulation by siRNA.

Table 4.2. Human ChIP primers for HSP genes.

Target gene (primer)[a]	Primer number	Primer sequences
SPG5 F	RN3511	5'-TTGCTTGTCCGGCGCACCAG -3'
SPG5 R	RN3512	5'-TCCCCTCGCCATCTGGGTC-3'
SPG6 F	RN3630	5'-CCCCTCTTCCTGCTCCTCC-3'
SPG6 R	RN3631	5'-CACCAGGCTCGACACGACG-3'
SPG7 F	RN3464	5'-TCAGGCAGCCACGAGGTAGAC-3'
SPG7 R	RN3465	5'-AGCAGCAGCACGGCCATGTTG-3'
SPG11 F	RN3622	5'-ACGAATGGAATCGACCGGAG-3'
SPG11 R	RN3623	5'-CTTCCTCTGCAGCCATCTTG-3'

[a] Abbreviations: F: forward PCR primer; R: reverse PCR primer. Gene sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/Homo_sapiens/Info/Index).

Table 4.3. ChIP primers for the *NRF1* promoter.

Target gene (primer)[a]	Primer number	Primer sequences
Human *NRF1*promoter F	RN3511	5'-TTGCTTGTCCGGCGCACCAG -3'
Human *NRF1* promoter R	RN3512	5'-TCCCCTCGCCATCTGGGTC-3'
Human *SNURF-SNRPN* enhancer F	RN2683	5'-CAGGTCTTGGAAGGCTATGTCG-3'
Human *SNURF-SNRPN* enhancer R	RN2684	5'-CTCCCCACTGGCGGCTCTAC-3'
Marsupial *Nrf1* promoter F	RN3464	5'-TCAGGCAGCCACGAGGTAGAC-3'
Marsupial *Nrf1* promoter R	RN3465	5'-AGCAGCAGCACGGCCATGTTG-3'
Marsupial *Snrpb* promoter F	RN3751	5'-AATAAGAAACAATGAATATGTGAGG-3'
Marsupial *Snrpb* promoter R	RN3752	5'-CACGACAAAACGGTCCGACTG-3'
Marsupial *Tfam* promoter F	RN3622	5'-ACGAATGGAATCGACCGGAG-3'
Marsupial *Tfam* promoter R	RN3623	5'-CTTCCTCTGCAGCCATCTTG-3'

[a] Abbreviations: F: forward PCR primer; R: reverse PCR primer. Gene sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/Homo_sapiens/Info/Index).

4.2.5 Reporter constructs and luciferase reporter assays

The minimal promoters of *SPG4* and *SPG6* were cloned by PCR into a pGL3-enhancer (pGL3e) luciferase vector using primers from **Table 4.4.** Either the pGL3e-*SPG4* or -*SPG6* promoter construct was transfected into mammalian cells with 1) control siRNA; 2) siRNA targeting luciferase mRNA; or 3) siRNA targeting *NRF1* mRNA.

Based on sequence alignments of conserved promoter sequence of *NRF1*, luciferase reporter constructs were made by PCR cloning using primers in **Table 4.5** to clone the minimal promoter sequence using genomic DNA from human SK-N-SH cells as template. PCR products were digested with *Bgl*II and *Hin*dIII for promoters or *Bam*HI and *Sal*I for enhancers, and the respective products ligated directionally into the pGL3basic vector (Promega). Site directed mutations were performed using primers in **Table 4.5** with the GeneTailor™ Site-Directed

Mutagenesis Kit (Invitrogen) according to manufacturer's instruction, and the mutated products were amplified by Platinum *Taq* DNA polymerase High Fidelity (Invitrogen). Each construct was sequenced to verify correct inserts.

To study the parameters that affect NRF1 binding ability, six minigenes were synthesized in pIDTSMART vectors (by Integrated DNA Technology, Inc.), each containing 10, 8, 6, 4, 2, or 1 copies of NRF1 consensus recognition sites (**Table 4.6**). These minigenes were subcloned from the pIDTSMART plasmids into the promoter position of a modified pGL3basic luciferase reporter vector (**Figure 4.6**).

SK-N-SH cells were grown in a 6-well cell culture plate at a concentration of 1.8×10^5 cells/well and each well cultured in 2 ml cell medium. When the cells reached 90% confluence, 1 μg of each construct was transfected along with the pRL vector (Promega), a second luciferase vector used for normalization in the Dual-Luciferase® Reporter Assay System (Promega), with Lipofectamine 2000 (Invitrogen) into the cells. 24 hr after transfection, cell medium was removed from the cell culture, and washed once with PBS. 500 μl passive lysis buffer (Promega) was applied to each well and the culture plate was rocked at room temperature for 15 min. The lysates were then collected into a 1.5 ml tube and 100 μl LARII (Promega; freshly prepared for each luciferase assay) was added. The first luciferase activity reading was measured on a 20/20n luminometer (Turner Biosystem). Subsequently, 100 μl Stop & Glo® reagent (Promega) was added which allows the measurement of a second luciferase activity (*Renilla*) reading to normalize the results. The luciferase assay was repeated independently three times to test statistical significance by a student's t-test.

Table 4.4. Primers for *SPG4* and *SPG6* promoter luciferase constructs.

Target gene (primer)[a]	Primer number	Primer sequences[b]
Human *Spastin* (*SPG4*) pr F	RN3470	5'-GCTT**ACGCGT**TGAGCCGAACTGCACATTGG-3'
Human *Spastin* (*SPG4*) pr R	RN3471	5'-CATA**CTCGAG**GTCTCAGGAGCTCCGCACTG-3'
Human *NIPA1* (*SPG6*) pr F	RN3472	5'-GCTT**ACGCGT**GCCCGCGCCTCCCGGTCACC-3'
Human *NIPA1* (*SPG6*) pr R	RN3473	5'-CATA**CTCGAG**GCAGCTGCAGTCCCCATTCC-3'

[a] Abbreviations: pr: promoter; F: forward PCR primer; R: reverse PCR primer. Gene sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/Homo_sapiens/Info/Index).

[b] Enzyme digestion sites are in bold and underlined.

Table 4.5. Primers for *NRF1* promoter luciferase constructs.

Target gene (primer)[a]	Primer number	Primer sequences
NRF1 sub-cloning primers into pGL3-basic promoter site F	RN3622	5'-TCCA**AGATCT**GACCTCCTCATCATGGCGG-3'
NRF1 sub-cloning primers into pGL3-basic promoter site R	RN3623	5'-AATT**AAGCTT**CTCAGAGCGGCTGCGCTAC-3'
NRF1 promoter YY1 mutation minigene for luciferase reporter assays	RN3846	5'-**AGATCT**gacctcctcatcatggcggcggccgggcggggaagccgggccgtgcgt cgcgtgcgtgccctctccctcccctcccctcctcggcggcggcggcggcggcagaa gcggcagcgct**cgcAaAAT**ccgctggtggcaggaggctgcgaggagccggcgcggtc gcagtctccacg**gcgcaggccca**cggta**gcgcagccgc**tctgag**AAGCTT**-3'
NRF1 promoter 5' NRF1-like mutation minigene for luciferase reporter assays	RN3847	5'-**AGATCT**gacctcctcatcatggcggcggccgggcggggaagccgggccgtgcgt cgcgtgcgtgccctctccctcccctcccctcctcggcggcggcggcggcggcagaa gcggcagcgct**cgccattg**ccgctggtggcaggaggctgcgaggagccggcgcggtc gcagtctccacg**gAATaggAcca**cggta**gcgcagccgc**tctgag**AAGCTT**-3'
NRF1 promoter 3' NRF1-like mutation minigene for luciferase reporter assays	RN3848	5'-**AGATCT**gacctcctcatcatggcggcggccgggcggggaagccgggccgtgcgt cgcgtgcgtgccctctccctcccctcccctcctcggcggcggcggcggcggcagaa gcggcagcgct**cgccattg**ccgctggtggcaggaggctgcgaggagccggcgcggtc gcagtctccacg**gcgcaggccca**cggta**gAATagccAc**tctgag**AAGCTT**-3'
NRF1 promoter 5' & 3' NRF1-like mutations minigene for luciferase reporter assays	RN3849	5'-**AGATCT**gacctcctcatcatggcggcggccgggcggggaagccgggccgtgcgt cgcgtgcgtgccctctccctcccctcccctcctcggcggcggcggcggcggcagaa gcggcagcgct**cgccattg**ccgctggtggcaggaggctgcgaggagccggcgcggtc gcagtctccacg**gAATaggAcca**cggta**gAATagccAc**tctgag**AAGCTT**-3'

[a] Abbreviations: F: forward PCR primer; R: reverse PCR primer. Gene sequences were obtained from the Ensembl genome browser (http://www.ensembl.org/Homo_sapiens/Info/Index).

[b] Enzyme digestion sites are in bold and underlined. YY1 site is in bold pink. NRF1 site is in bold red. Mutated nucleotides are in uppercase.

Table 4.6. Minigenes for NRF1 tandem arrays in luciferase constructs.

Minigenes with copy number[a]	Primer number	Primer sequences[b]
NRF1 site tandem array (10 copies) in a 230-bp mini gene	RN3755	5′-TTAA**CTCGAG**CGCGCATGCGCACCTCTGGGCC**GTCGCATGCGCA**CCAGTCTAGC**CG** **CGCAGGCGCG**CTAGTCAGGC**TGCGCATGCGCG**CCATTGCCTG**TGCGCAAGCGCG**CTACT CGCAC**TGCGCATGCGCA**CCTTTGCATG**GGCGCCTGCGCA**CAATGGTCTG**CGCGCAGGCG** **CA**CACGTCAGAC**TGCGCAGGCGTG**CACGTCGGGC**GGCGCATGCGCA****AGATCT**TGTG-3′
NRF1 site tandem array (8 copies) in a 186-bp mini gene	RN3756	5′-TTAA**CTCGAG**CGCGCATGCGCACCTCTGGGCC**GTCGCATGCGCA**CCAGTCTAGC**CG** **CGCAGGCGCG**CTAGTCAGGC**TGCGCATGCGCG**CCATTGCCTG**TGCGCAAGCGCG**CTACT CGCAC**TGCGCATGCGCA**CCTTTGCATG**GGCGCCTGCGCA**CAATGGTCTG**CGCGCAGGCG** **CA****AGATCT**TGTG-3′
NRF1 site tandem array (6 copies) in a 142-bp mini gene	RN3757	5′-TTAA**CTCGAG**CGCGCATGCGCACCTCTGGGCC**GTCGCATGCGCA**CCAGTCTAGC**CG** **CGCAGGCGCG**CTAGTCAGGC**TGCGCATGCGCG**CCATTGCCTG**TGCGCAAGCGCG**CTACT CGCAC**TGCGCATGCGCA****AGATCT**TGTG-3′
NRF1 site tandem array (4 copies) in a 98-bp mini gene	RN3758	5′-TTAA**CTCGAG**CGCGCATGCGCACCTCTGGGCC**GTCGCATGCGCA**CCAGTCTAGC**CG** **CGCAGGCGCG**CTAGTCAGGC**TGCGCATGCGCG****AGATCT**TGTG-3′
NRF1 site tandem array (2 copies) in a 54-bp mini gene	RN3759	5′-TTAA**CTCGAG**CGCGCATGCGCACCTCTGGGCC**GTCGCATGCGCA****AGATCT**TGTG- 3′
NRF1 site tandem array (1 copy) in a 32-bp mini gene	RN3760	5′-TTAA**CTCGAG**CGCGCATGCGCA****AGATCT**TGTG-3′

[a] NRF1 binding sequences are experimentally confirmed NRF1 binding sites or high affinity sites based on the consensus sequence: 5'-YGCGCANGCGCR-3'. DNase I footprint assay showed that at least 4 nucleotides on flanking sequence was needed for NRF1 binding (Virbasius & Scarpulla, 1994).

[b] NRF1 sites are in bold red. Enzyme digestion sites are in bold and underlined.

4.3 RESULTS

4.3.1 NRF1 regulates hereditary spastic paraplegia genes *in trans*

In this study, we have identified potential NRF1 binding sites in the promoters a total of 10 known HSP loci with putative NRF1 binding sites in the 5' regulatory region (**Table 4.7**). Importantly, the putative NRF1 binding sites in the promoters of most of the HSP loci were evolutionary conserved in all eutherian mammals, and for some sites in a marsupial (*M. domestica*) and a monotreme (platypus) (examples are shown in **Figure 4.2** for *SPG4*, *SPG5*, *SPG6* and *SPG7*). Evolutionary maintenance of so-called "phylogenetic footprints" provides strong evidence that a DNA-binding protein recognizes that sequence. In several of the HSP gene promoters, such as for *SPG6* and *SPG7*, putative NRF1 and Sp1 (a TF binding at most "CpG-island" of mammalian gene promoters) motifs were the only predicted conserved binding sites and for others, such as *SPG5* and *SPG4*, the putative NRF1 binding site(s) were the main conserved binding sites, attesting to the likely significance of NRF1 regulation.

To show that NRF1 physically binds *in vivo* at these regulatory elements, we used ChIP in a human neuroblastoma cell line, SK-N-SH, that robustly expresses mRNA for all investigated HSP genes (**Figure 4.3;** data not shown). The ChIP assay confirmed that NRF1 binds specifically and robustly to the promoters at all predicted sites for HSP genes in human SK-N-SH cells (examples are shown in **Figure 4.4**, summarized in **Table 4.5**). Similarly, ChIP assays were performed in mouse Neuro2A cells and NRF1 binds robustly to all predicted sites for mouse orthologs of HSP genes (data not shown). NRF1 regulation of HSP loci within neuronal-like cells was further confirmed by knocking down the steady-state level of endogenous *NRF1* mRNA using silencing RNA (siRNA) studies. The siRNA treatment was effective, knocking down the

NRF1 mRNA level dramatically but not affecting the mRNA level of a control gene, *GAPDH* (**Figure 4.5**). Indeed, all tested HSP loci having NRF1 motifs that bind the TF in ChIP assays demonstrated a major reduction in mRNA levels in the *NRF1* siRNA-treated cells (**Figure 4.5**).

Utilizing luciferase reporter vectors, further assays of transcriptional promoter function have been employed to molecularly dissect the transcriptional mechanisms for HSP gene promoters. Functional promoters will drive transcription, with the degree of light produced a marker of the amount of luciferase enzyme produced and hence of transcriptional activation levels. We did this for the evolutionary conserved segments of the *SPG4* and *SPG6* promoters (**Figure 4.6**), with high levels of promoter activity found [**Figure 4.6(b) & (c),** control siRNA columns]. Using co-transfection with a siRNA targeting the luciferase mRNA, promoter activity was significantly down-regulated [**Figure 4.6(b), (c)**], while both *SPG4* and *SPG6* promoter activity was essentially abolished using siRNA targeting the endogenous *NRF1* mRNA [**Figure 4.6(b), (c)**].These results confirm the phylogenetic predictions, that NRF1 provides the major regulatory activity at these promoters.

Table 4.7. Summary of HSP loci with NRF1 regulation.

HSP locus (Gene)	Human Chr[a]	Presence of putative NRF1 site by bioinformatics[b]	Function	ChIP confirmation in SK-N-SH cells
SPG4 (*SPASTIN*)	2p22	+	Cytosolic microtubule severing protein, with AAA and MIT domains	Yes
SPG5 (*CYP7B1*)	8q12.3	+ (intron 1)	Cytochrome p450 enzyme, cholesterol and steroid metabolism	Yes
SPG6 (*NIPA1*)	15q11.2	+	Transporter; in ER and endosome membranes	Yes
SPG7 (*PARAPLEGIN*)	16q24.3	+++	Mitochondrial protein, AAA ATPase domain	Yes
SPG8[c] (*STRUMPELLIN*)	8q24	+	Protein function unknown, contains spectrin domains, involved in motor-neuron outgrowth	Yes
SPG11 (*SPATACSIN*)	15q15.3-q21.1	L[a]	Not known (proposed TM domains not verified by our analyses)	Yes
SPG13 (*HSP60*)	2q24-q34	++L	Mitochondrial chaperonin, heat shock protein	Yes (proximal and distal)
SPG20 (*SPARTIN*)	13q12.3	+	MIT domain, endosomal protein	Yes
SPG33 (*ZFYVE27*)	10q24.2	+	Promotes neurite formation, involved in membrane traffic events	Yes
SPG42 (*SLC22A1*)	3q25.31	_[d]	Acetyl-CoA transporter	Not tested

[a] Abbreviations: chr: chromosome; L: NRF1-like site (sites with one mismatch in one GCGC motif are defined as NRF1 like site).

[b] All confirmation of NRF1 sites was by ChIP analysis using human SK-N-SH and mouse Neuro2A neuroblastoma cell lines, and for *SPG4* and *SPG6* also by luciferase reporter promoter analyses. In addition, NRF1 regulation of all these genes was seen using siRNA that targeted and knocked down the level of *NRF1* mRNA in human SK-N-SH neuroblastoma cells.

[c] No NRF1 site occurs in the 5' regulatory region in rodents; the NRF1 site in human *SPG8* and NRF1-like site in *SPG11* are conserved in all other non-rodent sequenced mammalian genomes.

[d] No NRF1 site in the 5' regulatory region in primates; a site in mouse *Spg42* is conserved in all other sequenced mammalian genomes.

(a) CLUSTAL 2.0.10 multiple sequence alignment for the conserved part of the *SPG4* (*SPASTIN*) promoter

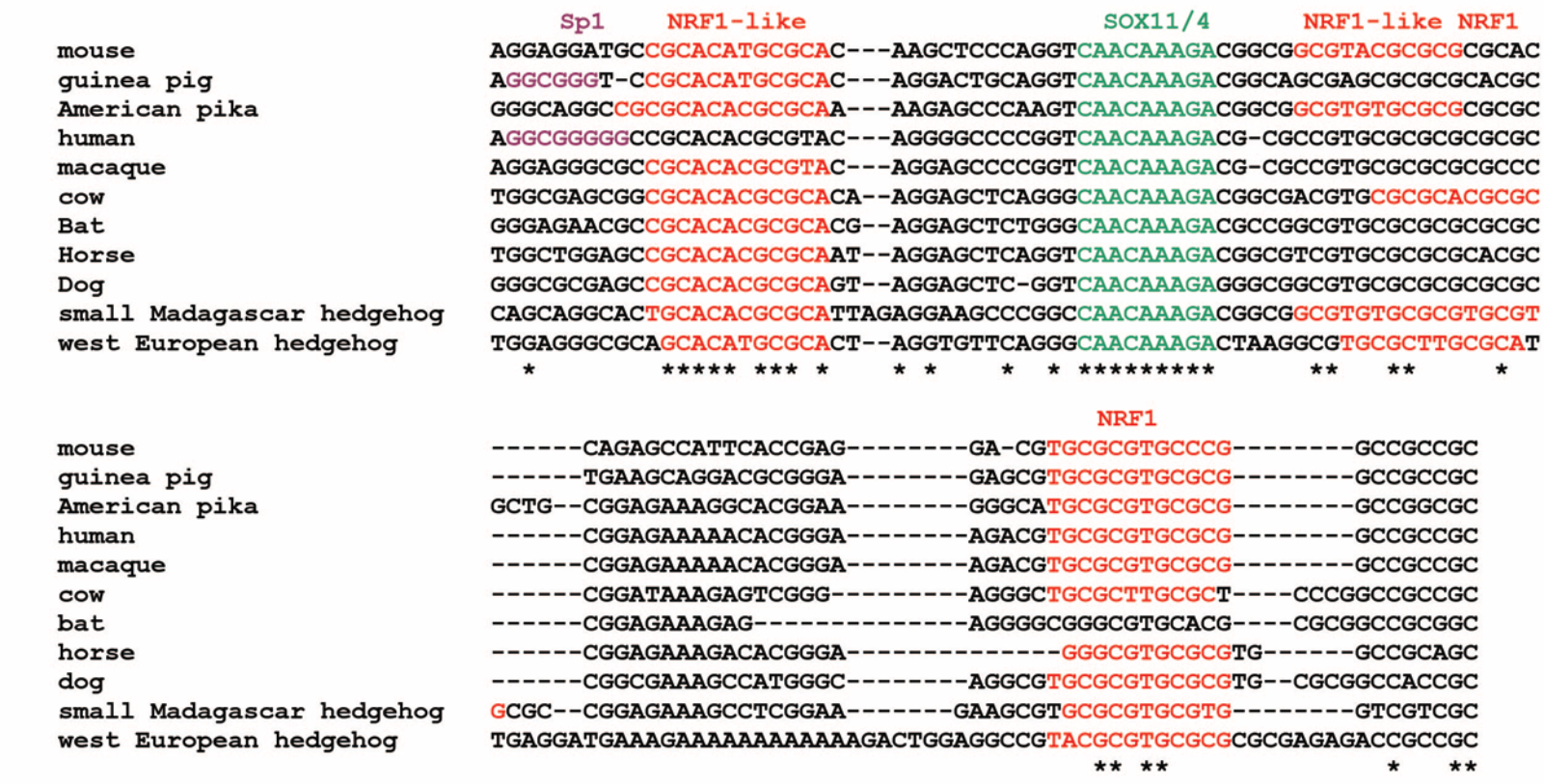


Figure 4.2. (a) (continued below)

(b) CLUSTAL 2.0.8 multiple sequence alignment for the *SPG5* (*CYP7B1*) exon 1-intron 1 region

```
                    NRF1              Exon 1  Intron 1                                                    NRF1
armadillo           CCTGTGCCTGCGCGCCCGCAGGACCAGGTAAGCGCC--GGCGGCCCGAGGGCG--GCCCCGTAG------CCGGGCCACGCATGCGCCCTCGC-CCC-GCGGACCGCGCGCGCGACCG
cow                 CTTGTGCCTGAATGCGCGGCGCACCAGGTAAGCGCCCCGACGACCCCGCGCCA--GCCC------------GGGCCACGCATGCGTCCTGGG-CCC---AGACCGCCCTCCATCC
horse               CCTGTGTCTGCGCGCCCAGCGCACCAGGTAAGCACCCCAGCCGCCTCGCCGCGAGGCGCAGCGGA-----CCGGGCCACGCATGCGCCCTGGG-CCCTGCGGACCGCGCGCGCGACCC
human               CCTCTGCTTGCTTGTCCGGCGCACCAGGTAAGCGCCTCGGCCGCC------------------------AGGCCACGCATGCGCCCTGGG-CCCCGCGGGCCGCGCGCGACCC
American pika       CCTGTGCCTGTGTGGCCGGCGCACCAGGTAAGCGCCCCGCCGCCC--------------------------GGACACGCATGCGCCCTGGG-CCCCGCCGCC---GTGCGACCC
dog                 CCTGTGCCTGCGCGCCCGGCGCACCAGGTAAGCGCCCCCGCGGCCTCGCCGCGCGGCCGGGCTC------CCGGGCCACGCATGCGCCCTGGT-CCCCGCGGGCCGAGCGCGACCC
cat                 CTTGTGCCTGACCTCCAGGCGCACCAGGTAAGCGCCCCCGCGACCTCACCGCGCGGCCGGGCA-----GCCAGGGCCACGCATGCGCTCTGGG-CCCCGCGGACCGCGCGCGACCC
European shrew      CCAGTGCCTGCGCTCCCGGCGCACCAGGTAAGTTTCCCCGCGGCCCCGCGCGTCGGCCGCGCGCCACGGCCCGGGCCACGCATGCGCTCTGGG-CGTCCCCGTCGGCTCGCGACCT
tree shrew          GCTGTGCCTGAGCTCCAGGCGCACCAGGTAAGCGCCCAGGCCGCC--------------------------GGGGCCACGCATGCGCTCTGGG-CCCTGCGGGCCGCGCGCGACCC
rabbit              CCTGTGCCTGCGCGGCCGGCGCACCAGGTAAGCGCCCCTGCCGCCGC------------------------GGCCACGCATGCGTCCTGGG-CCCCGCGGCC---GCGCGACCC
lemur               CCTGTGCCTGCGCGCACGGCGCACAAGGTAAGCGCCCCGACCGCCG-------------------------GGCCACGCATGCGCCCTGAAACCCCGGGGCCCGCGCGCGACCC
squirrel            CCTGTGCCTGCGCTCCAGGCGCATCAGGTAAGCGCCCGGGCCTTTG-------------------------GGGACACGCATGCGCCTTGGGCCCCCGCGGCCCGCGCGCGACCC
guinea pig          CCTGTGCCTGCGCGCCCGGCGCTCCAGGTAA---CCCCGGCGGGC--------------------------CCACGCATGCGCCCTGG---CCCGCGGCC---GCGCGCCCC
mouse               CCTGTTCCTCCTTACCCGGCGCACCAGGTAAGCGCCT-GGCT-----------------------------GGCCACGCATGCGT---GGG--ATCGCGGCC---ACGCGACCT
rat                 CCTGTTCCTCCTCACCCGGCGCACAAGGTAAGCGCCTAGGCT-----------------------------GGCCACGCATGCGT---GGG--ATCGCGGCC---ACGTAACCT
                        *   *           *     *****    *    *                                  **********                 *         *
```

(c) CLUSTAL 2.0.8 multiple sequence alignment for the *SPG6* (*NIPA1*) promoter

```
                                  NRF1    NRF1          Sp1         Sp1
human                   GGGGC---GCGGCGCGCAGGCGCAGGCTCGGAGGGC-GGGCGCGGGCG------------GAATGGGGACTGCAGCTGCGGCAGCG
baboon                  CGGGC---GCGGCGCGCAGGCGCAGGCTCGGAGGGC-GGGCGCGGGCG------------GAATGGGGACTGCAGCTGCGGCAGCA
rhesus monkey           CGGGC---GCGGCGCGCAGGCGCAGGCTCGGAGGGC-GGGCGCGGGCG------------GAATGGGGACTGCAGCTGCGGCAGCG
rabbit                  CGGGC---GCGGCGCGCAGGCGCAGGCACGGAGGGC-GGGCGCGGGCG------------GAATGGGGACTGCAGCGGCGGCAGCG
dog                     CCGGC---TCGGCGCGCAGGCGCAGGCGCGGAGGGC-GGGCGCGGGCGCGGGCGCGGGCGGAATGGGGACTGCAGCGGCGGCGGCG
west European hedgehog  CGGGC--GC-GGTGCGCAGGCGCGGA---GGGCGGC--GGCGCGGGCG------------GAATGGGGACTGCAGCGGGGGCAGCG
horse                   CGGGC--GCTGGCGTGCAGGCGCAGACAAGGACGGGAGGGCGCGGGCC-----------AGAATGGGGTCTGCAGCGGCGGCAGCG
cow                     CGGGCTGGCGAGCGCGCAGGCGCAGA---GGGCGGG----CACGGGCG------------GAATGGGGACTGCGGCTGCTGCAGCG
mouse                   GGGGCGCG-GCGCGCGCACGCGCAGC---GGAGGGC--GGCGCGGGCG------------GCATGGGGACTGCGGCGGCGGCTGCG
guinea pig              CGG-CGCA-GAG-GCGCAGGCGCGAGGGCG----------CGCGGGCG------------GCATGGGGACTGCAGCGGCTGCGGCG
European shrew          CGGGCT-TCTT-TTCGCAGGCGCAGGCGCGGGCCGC-----GCGGGCG------------GGATGGGGACTGCGGCGGCGGCAGCG
Monodelphis             CGTGAT--CCCAGCGGTGTGCGCAGGCGCGGGAACCCGGGCCGCTz------AGGGAGGAAGATGATGACTGCTACTGCAGCAGCC
platypus                CGTC-TCGGAGGGCCTTGCGCGCATGCGCCGG----CGGAGGAGGZ--------GGAGCGGGATGGCTGCAGTGGCCGGGGAAGGG
                                        ****                                          ***    * *   * *  *  *
```

where z (36-nt) and Z (28-nt) are insertions in *M. domesticus* and platypus, respectively:

```
Monodelphis     TTCGGTAGCAAGTAGGGAGGAAGCTTCGGCTCTGCT
platypus        TTCGGGAGC-----GGGAG-CCGGAGCGG--GAGCC
                ***** ***     *****   *   ***    **
```

Figure 4.2. (b), (c) (continued below)

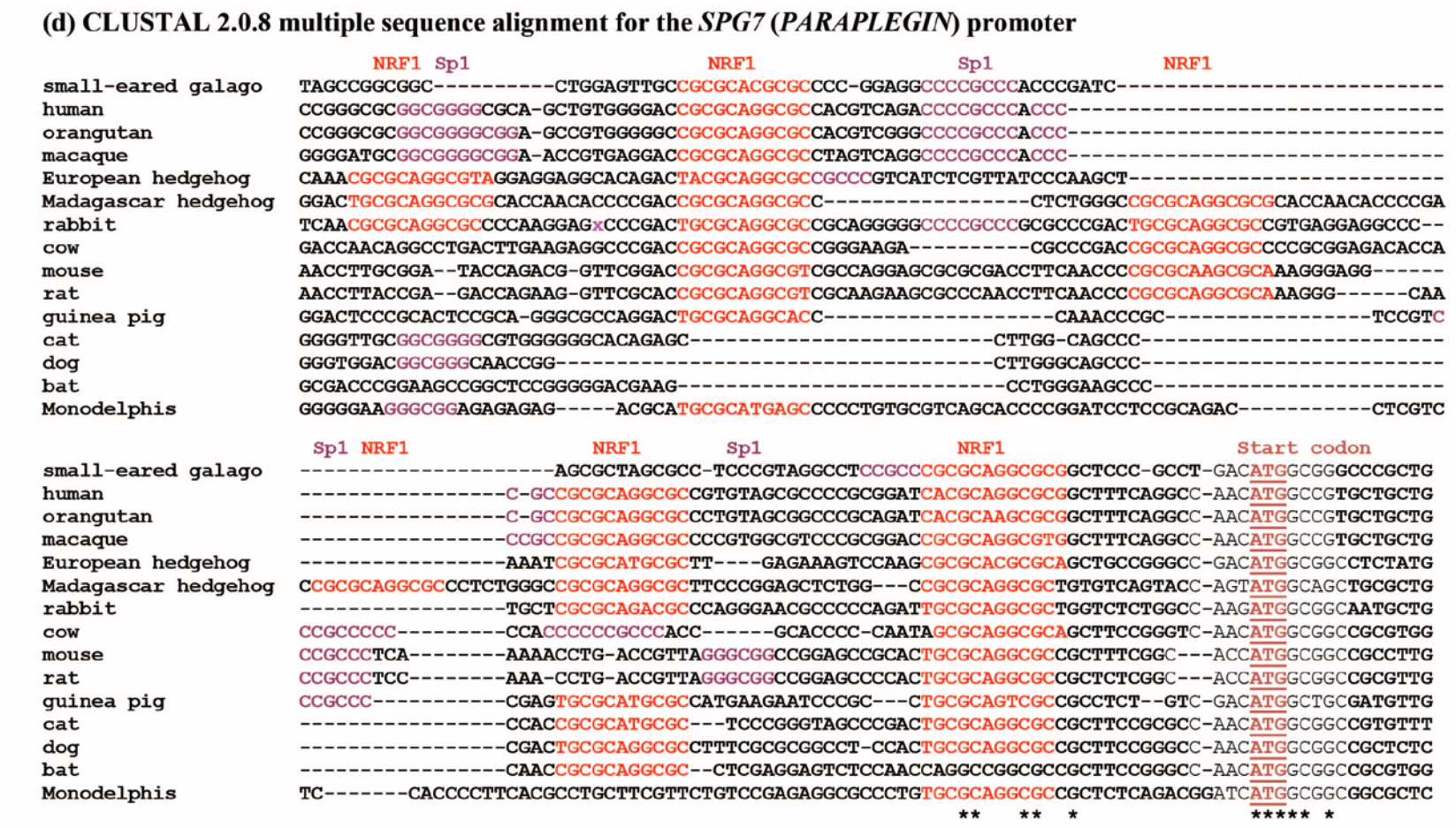


Figure 4.2. Evolutionary conserved NRF1 binding motifs in 5' regulatory regions of HSP loci. Multi-sequence alignments were generated using Clustal W with manual adjustments, and representative examples are shown (* = conserved nucleotides): (a) *SPG4*, (b) *SPG5*, (c) *SPG6*, and (d) *SPG7*. In (c), z (36-nt) and Z (28-nt) are species-specific insertions in *M. domestica* and platypus, respectively; in (d), x is a 31-nt insertion in the rabbit.

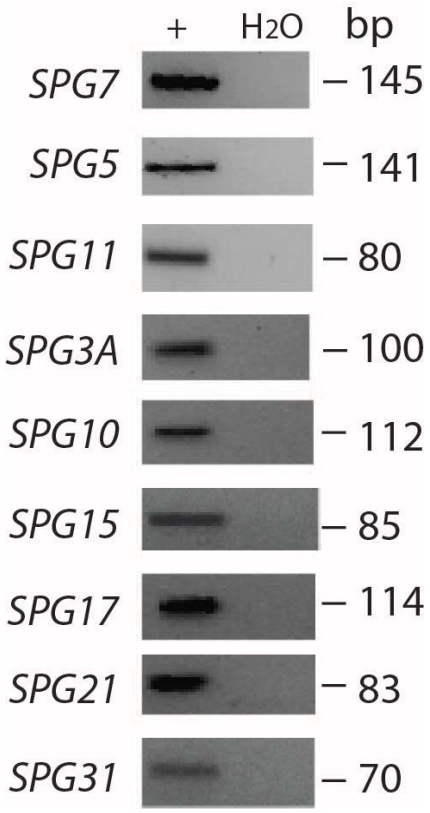


Figure 4.3. HSP gene expression studies in human SK-N-SH cells. bp: base pair. H_2O is PCR control.

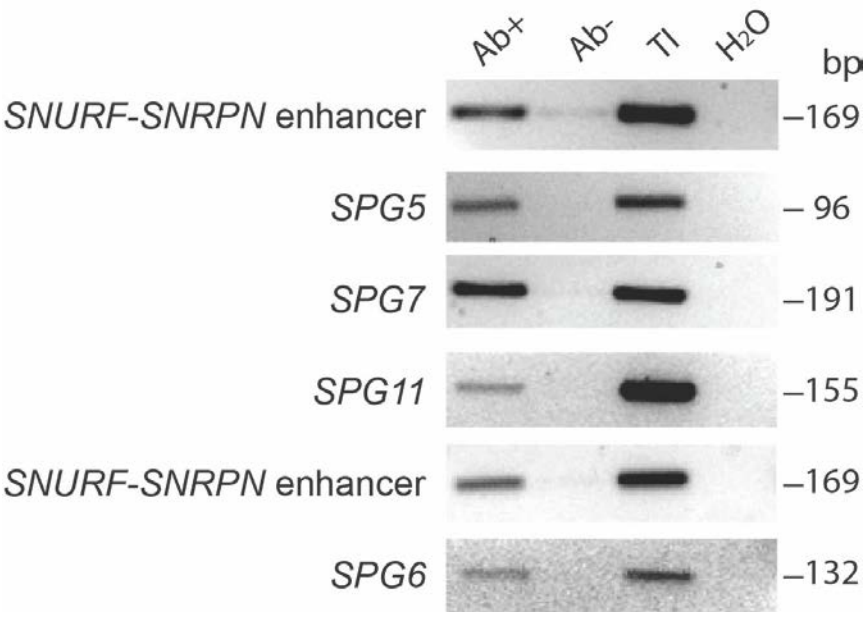


Figure 4.4. NRF1 binds at HSP gene promoters in human SK-N-SH cells by ChIP. Antibody positive is shown as Ab+, controls are no antibody (Ab-), total input (TI) DNA is for PCR positive control, H_2O is PCR negative control, and the positive ChIP control is the *SNURF-SNRPN* enhancer (Rodriguez-Jato *et al.*, 2005).

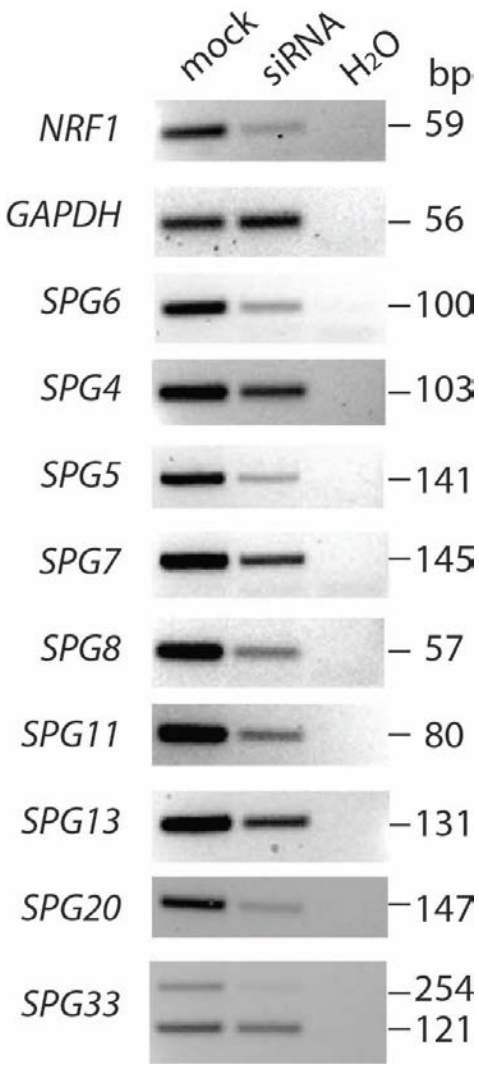


Figure 4.5. siRNA targeting NRF1 mRNA knocks down HSP gene expression. SK-N-SH cells were transfected with pSUPER-*NRF1* siRNA or mock transfected as a control, with qualitative analysis of gene expression including a negative control (*GAPDH*). The *GAPDH* mRNA level is unaffected by *NRF1* siRNA, whereas the mRNA levels for *NRF1* and all shown HSP genes are reduced by the siRNA treatment (including two mRNA splice isoforms for *SPG33*).

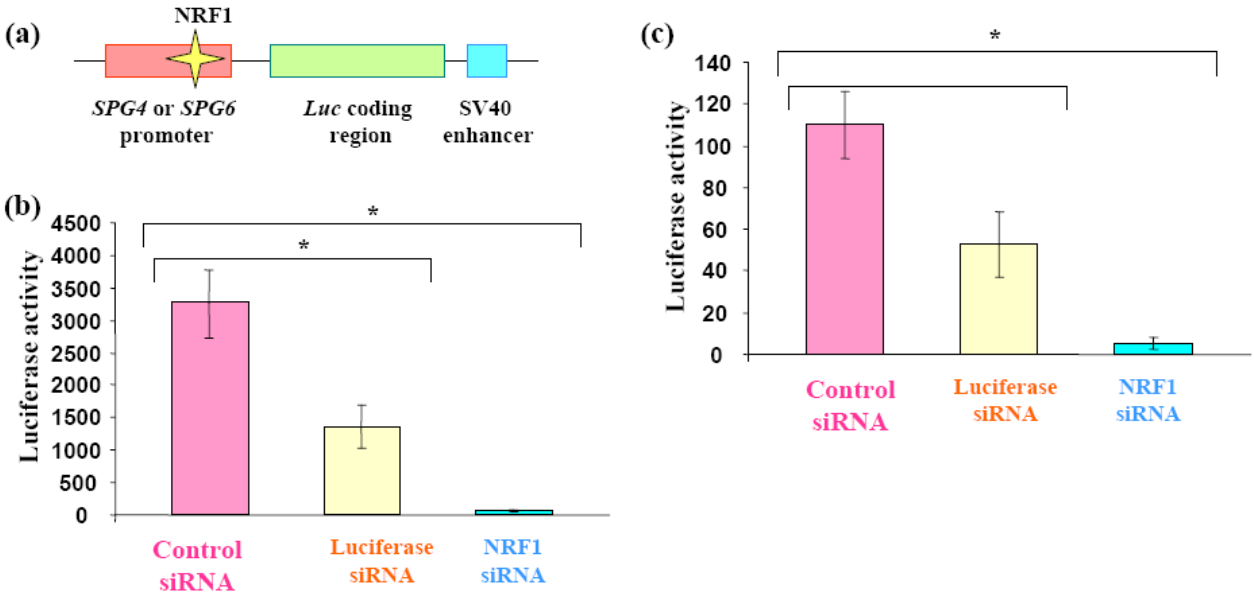


Figure 4.6. siRNA targeting the NRF1 mRNA ablates *SPG4* and *SPG6* promoter function. A pGL3e-SPGx-promoter [x = 4 or 6; constructs are shown in **(a)**] plasmid and a siRNA plasmid (either targeting an *Arl2* control, *Luc*, or *NRF1*) were co-transfected into SK-N-SH cells. Luciferase activities were measured for the **(b)** *SPG4* promoter and **(c)** *SPG6* promoter. *, $P < 0.05$, for the control siRNA versus test siRNA (targeting *Luc* or *NRF1*).

4.3.2 NRF1 autoregulates and enhances transcriptional activity

Since we found evolutionary conserved potential NRF1 binding sites in the 5' regulatory region of the gene, we then used molecular assays to confirm these findings. First, to test whether NRF1 binds this region, we used a ChIP assay and confirmed that NRF1 binds this site in the human SK-N-SH neuroblastoma cell line [**Figure 4.7(a)**], in the mouse Neuro2A neuroblastoma cell line (data not shown), and to the conserved, homologous region in a marsupial fibroblast cell

line [**Figure 4.7(b)**]. As shown by phylogenetically conserved motifs (**Figure 4.1**), YY1 is another conserved binding motif we found in the promoter of NRF1. Similarly, we used ChIP to confirm that YY1 binds to the *NRF1* promoter in SK-N-SH cells [**Figure 4.7(c)**] and Neuro2A cells (data not shown).

Next, we used luciferase reporter assays to test the degree of effects that each TF has on the regulation of NRF1. We cloned a minimal *NRF1* promoter-exon 1 sequence, including one Sp1 site, two NRF1 sites and one YY1 site, into a pGL3basic luciferase vector and generated mutations of 1) each and 2) both NRF1 sites, as well as 3) the YY1 site (**Figure 4.8**). Then, we co-transfected each pGL3basic derived construct with a *Renilla* luciferase vector as an internal control into a human peripheral neuronal precursor cell line, Flp-In 293 cells. As shown in **Figure 4.8**, the WT *NRF1* promoter-exon 1 promoter significantly increased luciferase activity by 87-fold compared to the empty vector ($P<0.001$), indicating a functional promoter construct. The promoter activities of all mutant constructs had a dramatic decrease in luciferase activity when compared to the WT construct (**Figure 4.8**). Specifically, when the YY1 site was mutated, the luciferase activity decreased by 67% or 29-fold ($P<0.05$). Mutation of either 5'- or 3'-NRF1 site decreased luciferase activity to 58% or 51-fold ($P<0.05$) and to 55% or 48-fold ($P<0.05$), respectively; whereas mutations of both NRF1 sites decreased luciferase activity to 44% or 39-fold ($P<0.05$). These results indicate that both NRF1 sites have a significant effect on regulating its own promoter to increase transcriptional activity. Nevertheless, the two NRF1 sites do not synergize in transcriptional activity, suggesting that only one NRF1 homodimer binds this region in its own promoter. This is probably due to the spacing between the two NRF1 sites in this region that is too small (4 nucleotides apart, **Figure 4.1**) for two NRF1 homodimers to bind

efficiently, since a minimum of 8 nucleotides between two NRF1 sites are needed (Virbasius & Scarpulla, 1994).

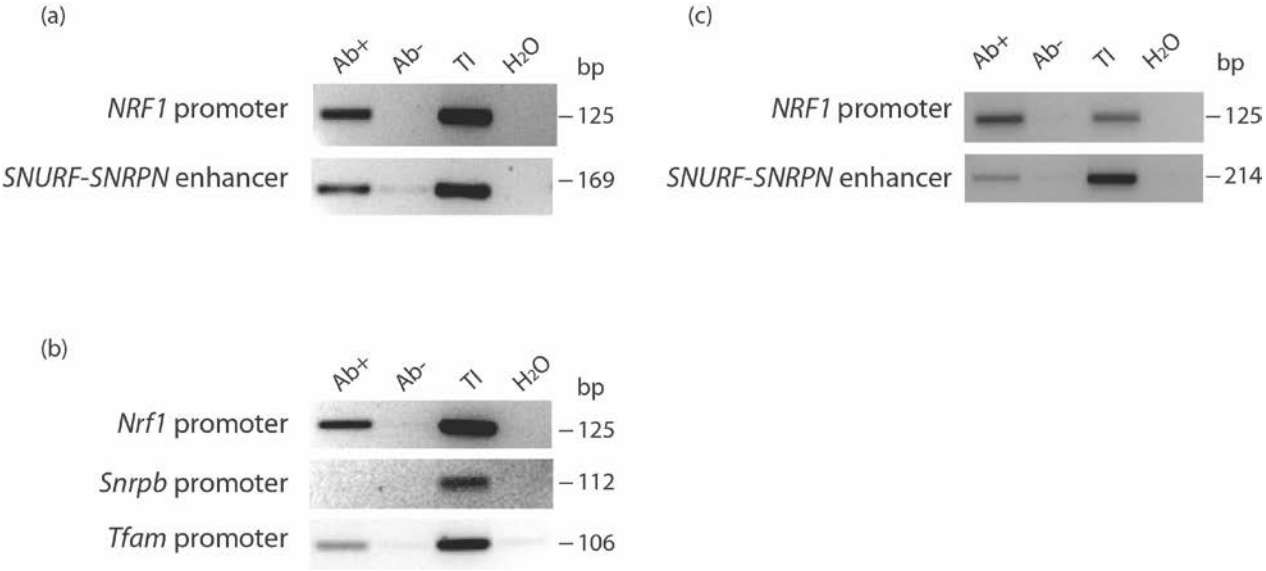


Figure 4.7. ChIP studies of the *NRF1* promoter-exon 1 region. (a) NRF1 ChIP in SK-N-SH cells. **(b)** NRF1 ChIP in marsupial cells. **(c)** YY1 ChIP in SK-N-SH cells. Antibody positive is Ab+, controls are no antibody (Ab), total input (TI) DNA is for PCR positive control, H_2O is PCR negative control, and the positive ChIP control is the *SNURF-SNRPN* enhancer (Rodriguez-Jato *et al.*, 2005).

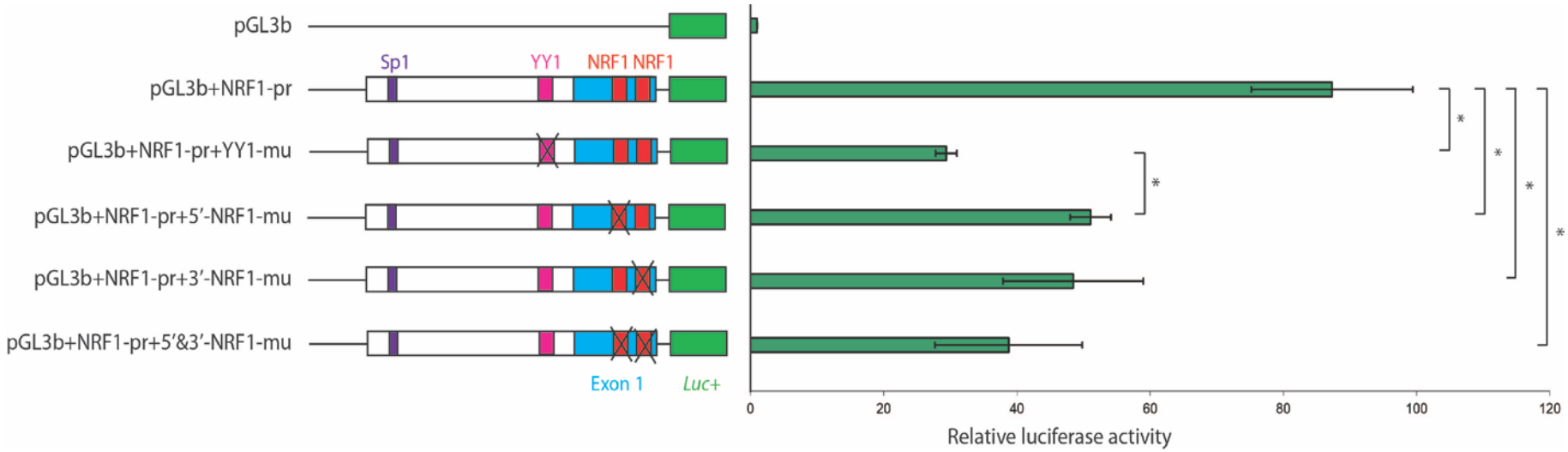


Figure 4.8. Luciferase reporter assays for *NRF1* promoter activity. *NRF1* promoter (pr) luciferase constructs (left) showing TF binding sites (Sp1, YY1, NRF1) and mutations (mu; X) in TF sites. Luciferase reporters assays (right). *, $P < 0.05$

4.3.3 Relationship between number of NRF1 sites and promoter activity

Luciferase reporter vectors with 10, 8, 6, 4, 2, or 1 copies of NRF1 binding sites were cloned into a modified pGL3basic vector to test how the number of NRF1 sites affect the level of transcriptional activity. The results indicated that these artificial NRF1-only sequences provided significant promoter function, and that luciferase activity increased as the number of NRF1 sites was increased [**Figure 4.9(a)**]. However, the relationship between the number of NRF1 sites in the promoter and effect on activating transcription was not purely linear [**Figure 4.9(a)**]. In this system, two NRF1 sites had 6-fold increased luciferase activity compared to one NRF1 site, while doubling to four NRF1 sites increased activity a further 14-fold (80-fold over a single NRF1 site). For unknown reasons, increasing to six or eight NRF1 sites had only no or small effect on luciferase activity, respectively, whereas ten NRF1 sites had a 2-fold further increase in promoter activity compared to eight NRF1 sites and for a total of 245-fold that of a single NRF1 site [**Figure 4.9(a)**]. In testing the correlation between the number of NRF1 sites versus the level of transcriptional activity, there was a strong correlation [correlation=0.94, **Figure 4.9(b)**]. In Excel, such data was plotted and fitted the best to a polynomial regression with order of 2 [R^2=0.9224, **Figure 4.9(b)**]. Therefore, I conclude that the number of NRF1 sites has positive effect in activating transcriptional activity of a promoter in the luciferase reporter system.

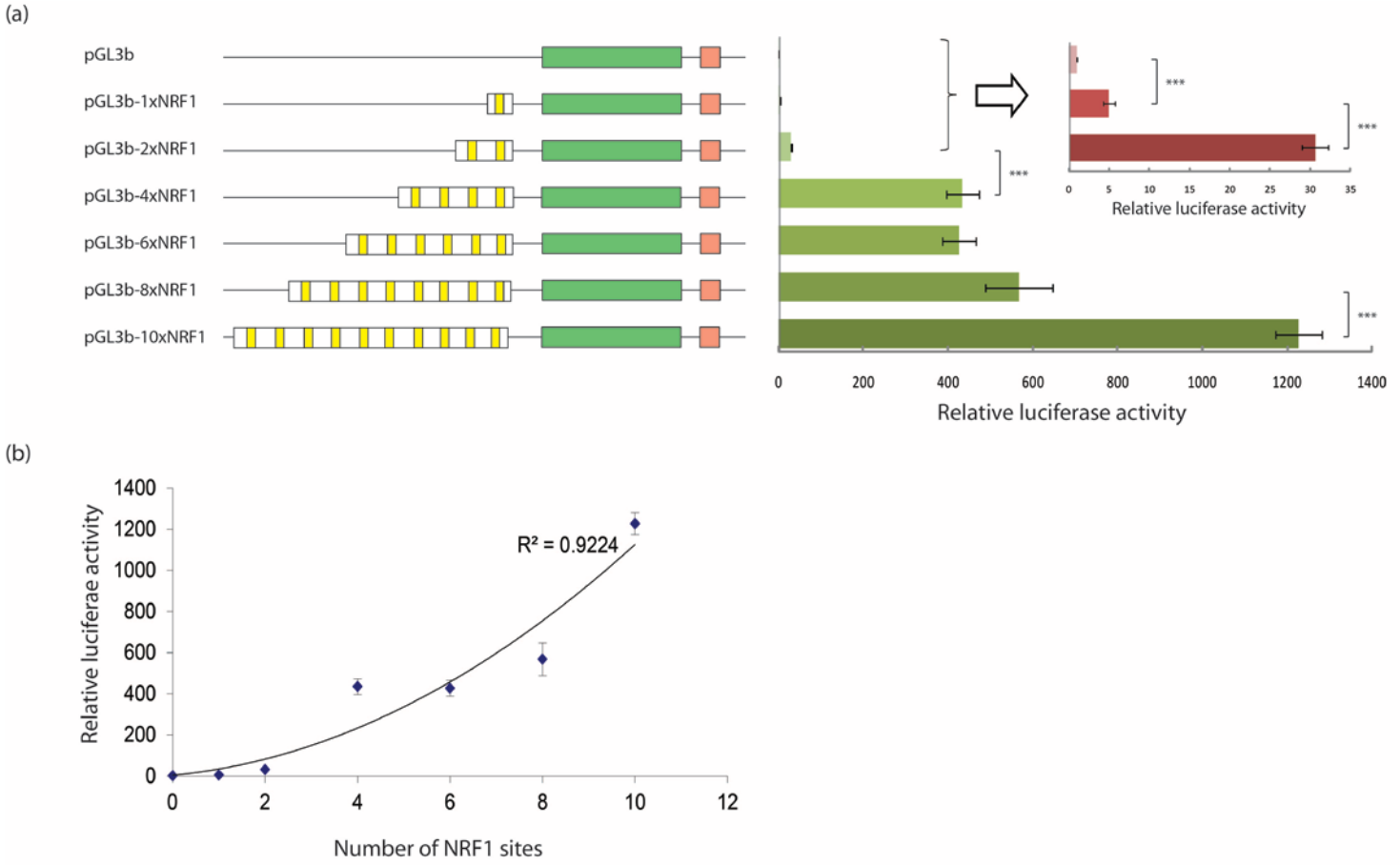


Figure 4.9. Luciferase reporter assays for NRF1 binding site tandem arrays. (a) NRF1 tandem array luciferase constructs (left). Luciferase reporters assays (right). *, $P < 0.05$; **, $P < 0.001$; *, $P < 0.0001$. (b) Correlation between relative luciferase activity versus the number of NRF1 sites.

4.4 DISCUSSION

4.4.1 NRF1 regulates spastic paraplegia genes *in trans*

By bioinformatics analysis and molecular assays, we have shown that NRF1 regulates many HSP genes. Our findings suggest that NRF1 may contribute significantly to gene regulation in the cortical neurons that extend corticospinal axons which degenerate in HSP. As would be expected, in some of these HSP gene promoters, there are also other conserved TF binding sites. For example, there is a highly conserved site for SOX4 and SOX11 in the *SPG4* promoter [**Figure 4.2(a)**]. Mutation of this site significantly reduced luciferase activity while activity of the *SPG4* promoter was significantly increased with SOX11 overexpression (B. Henson, R.D. Nicholls, personal communication).

As well as transcriptional regulation, it has become increasingly clear that a major additional regulatory mechanism exists in cells with modulation of mRNA stability and/or the level of translation of an mRNA by a novel functional class of very small RNA molecules termed microRNAs (miRNAs; Rivera & Bennett, 2010). Typically, mature, processed miRNAs are single-stranded RNAs 20-25 nucleotides in length (Chi *et al*., 2009). Mechanistically, miRNAs base-pair with their RNA targets, beginning with a 6-8 nucleotide "seed" sequence localized at the 5' end of a miRNA (Chi *et al*., 2009). The "seed" provides sufficient specificity for each miRNA to potentially regulate several hundred mRNAs with a range of less than 100 to over 1000 targets (Chi *et al*., 2009). Other members in our laboratory have shown that the miR-183/96/182 family miRNAs target sites in one-third of HSP mRNAs (B. Henson, R.D. Nicholls, personal communication). This provides further insight into the GRNs for HSP genes. An understanding of transcriptional plus miRNA regulatory signatures may help to identify and rank

the top candidate genes within large chromosomal domains linked to more than 20 uncloned HSP chromosomal regions (Blackstone *et al*., 2011).

4.4.2 NRF1 autoregulation

By phylogenetic analysis, we showed that NRF1 and YY1 binding sites in the *NRF1* promoter are within a functionally selected region over the past ~450 million years (**Figure 4.1**). These NRF1 and YY1 sites are also conserved in *Ciona* species (data not shown), among the most primitive vertebrates; further attesting to the importance of these regulatory features. Our molecular results demonstrated that NRF1 acts by a feed-forward mechanism to regulate its own transcriptional activity in a positive way. Based on the results from luciferase reporter assays, YY1 is also a transcriptional activator for the *NRF1* promoter. Since Sp1 is a common transcriptional activator in most mammalian promoters of the CpG-rich class (Pugh & Tjian, 1990), we hypothesize that there must be a negative mechanism of TF regulation associated with the *NRF1* promoter to repress its gene expression. Within intron 1 of *NRF1*, a conserved DNA element corresponds to a putative C/EBP (CCAAT/enhancer-binding protein) site (**Figure 4.1**). It was reported that C/EBP can either be a transcriptional activator (Landschulz *et al*., 1988) or repressor (Pei & Shih, 1990). Therefore, the next step in better understanding the transcriptional regulation of the *NRF1* gene is to test if C/EBP is the TF that binds to this site and whether this site positively or negatively regulates *NRF1* transcription. To further assess the complex question of how NRF1 regulates itself *in vivo*, it will be interesting to overexpress NRF1 protein in a cell line model and investigate how the increase of NRF1 protein level affects the transcription of *NRF1* mRNA to maintain homeostasis of the NRF1 protein levels.

4.4.3 Relationship between number of NRF1 sites and promoter activity

Here, I reported that number of NRF1 sites in an artificial promoter has a positive role in regulating transcriptional activity. Indeed, similar natural clusters of NRF1 sites occur in transcriptional regulatory sequences. For example, *SPG7* promoter has 1 (bat) to 6 (small Madagascar hedgehog) NRF1 sites [**Figure 4.2(d)**], *SPG4* promoter has up to 3 NRF1 sites [**Figure 4.2(a)**], "NRF1 cluster" enhancer in the PWS domain has 4 sites (**Figure 2.2**), and 5' regulatory regions of other candidate genes have various number of NRF1 sites, e.g., *EIF5A* (eukaryotic translation initiation factor 5A, 7 NRF1 sites), *MAP2K7* (mitogen-activated protein kinase kinase 7, 4 NRF1 sites), etc (data not shown).

Numerous studies have found that spacing between TF binding sites is also a critical determinant of gene expression (Vilen *et al*., 1991; Nolte *et al*., 2006; Cheung *et al*., 2007; Larsson *et al*., 2007). Therefore, it provides another direction in exploring the structure and affinity of NRF1 binding. The artificial NRF1 site arrays discussed above or the "NRF1 cluster" enhancer element from the PWS domain would provide suitable reagents to alter the spacing of the NRF1 sites to assess the effect on promoter or enhancer activity in *in vitro* luciferase studies. Similarly, the parameters of match to consensus NRF1 recognition sequence could also be studied in this assay system, since the structure of NRF1 has not been determined, such studies could provide insights into the properties of DNA binding by NRF1.

This study has served as a prelude to a future goal of determining the structure of NRF1, the mode of binding to target sites, and changes in affinity for different targets. In the long term, such studies may help to develop small molecule regulators of NRF1 function as drugs in diseases involving NRF1 deregulation, including diabetes, obesity, cancer and neurological diseases.

5.0 CONCLUSIONS AND FUTURE STUDIES

5.1 CONCLUSIONS

A GRN is a central processing unit in a living cell. GRN are composed of a collection of DNA segments which interact with each other by RNAs or proteins through intracellular and/or extracellular signals, thereby governing the rates at which genes in the network are transcribed into mRNA (Macneil & Walhout, 2011). A major class of regulators in a GRN is TF (Macneil & Walhout, 2011). The binding of TFs to specific DNA elements provides a major mechanism in the fundamental basis of a GRN (MacQuarrie *et al*., 2011). Other components in a GRN include RNA regulators, e.g., miRNA and lincRNA (long intergenic noncoding RNA, Wang *et al*., 2011), and chromatin factors, such as histone modifications (Macneil & Walhout, 2011). For over half a century, studies have been focused on TF regulation in order to study GRN (Chen & Rajewsky, 2007). The development of computational analysis in biological fields enables a preliminary scan for TF binding sites in the human genome, for which the biological importance can initially be tested by molecular phylogenetic analysis due to the wealth of animal whole genome sequences. The advent of new molecular techniques, such as ChIP–chip and ChIP-seq, allows more efficient direct experimental confirmation of bioinformatics results (Pareek *et al*., 2011). Other technologies, such as siRNA knockdown of an mRNA encoding a TF and hence of protein levels, or TF overexpression, followed by whole genome RNA analyses by expression

microarrays or RNA-seq (a next generation sequencing approach) also provide confirmation by dysregulation of expression of target genes for a specific TF. High throughput technologies in experimental model systems are now allowing assessment of more than one factor at a time in system approaches (Pareek *et al*., 2011).

In my research projects, I tried to make good use of both bioinformatics analysis and experimental techniques to elucidate the mechanisms and functional roles of NRF1 binding sites in the human genome. In particular, I have focused on NRF1 regulation in the 2 Mb PWS domain and the circadian system. In addition, I started to explore the function of NRF1 in regulating HSP genes and its own autoregulation. NRF1 binding sites were found in 5-6% of the promoters in the human genome by bioinformatics searches and ChIP-chip studies (Cam *et al*., 2004; FitzGerald *et al.*, 2004; Xi *et al*., 2007). It is known that many nuclear genes encoding mitochondrial products are regulated by NRF1 (Scarpulla, 2008). In our laboratory, bioinformatics analysis and molecular studies showed that NRF1 is overrepresented in the three systems we studied. NRF regulates 83% of the genes in the 2 Mb PWS region, 56% in circadian regulatory genes, and 45% in HSP genes (**Figure 5.1**). My studies have provided important evidence on NRF1 binding elements and initial studies towards identifying GRNs that involve NRF1.

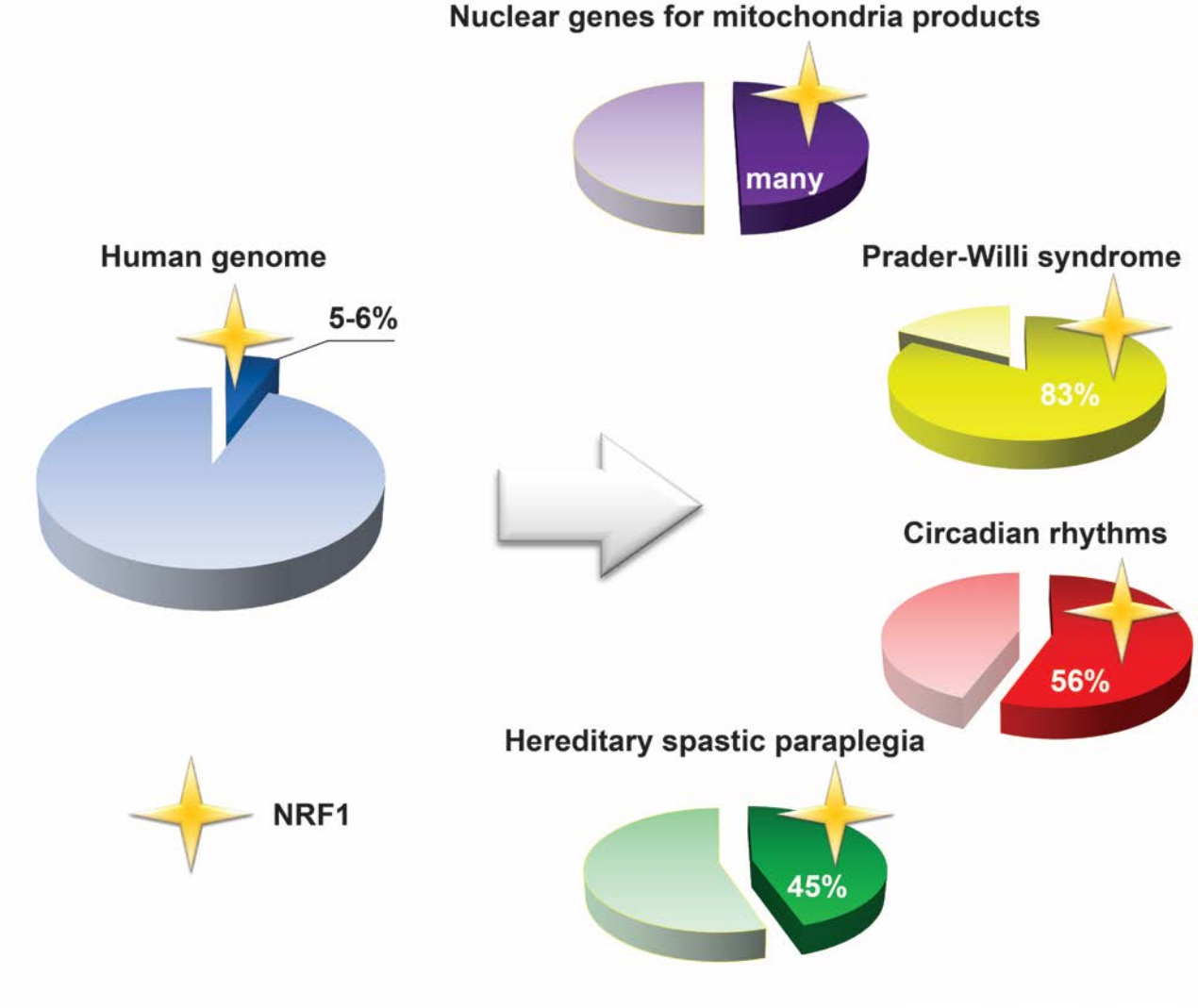


Figure 5.1. Systems with enrichment of transcriptional regulation by NRF1 in the human genome.

5.2 FUTURE STUDIES

For future work, we should not only focus on the study of specific TF binding sites such as NRF1, but also on *cis*-regulatory modules, that include binding sites for NRF1 and other TFs with which it may co-regulate different classes of genes. In **Chapter 2**, an original goal was to use somatic cell hybrids for a gene targeting experiment in order to study the mechanism of the "NRF1 cluster", a putative *cis*-regulatory module with enhancer function, on long-range regulation of the 2 Mb PWS-imprinted domain. Although my studies in somatic hybrid cells showed that none of these cell lines were suitable for gene targeting of the "NRF1 cluster" due to a lack of active chromatin at this site in these cells, the somatic cell hybrids studies nevertheless provided gene expression and chromatin configuration profiles of NRF1 binding and histone methylation for understanding gene regulation over the 2 Mb imprinted PWS domain. Optimally, the *cis*-regulatory role(s) of the "NRF1 cluster" can be investigated by gene targeting in mouse embryonic stem cells and subsequent analysis in knockout mice, to determine the extent to which this element is required for active chromatin an how many of the genes it regulates in a neural-specific manner within the PWS-orthologous domain in the mouse.

In the study of circadian rhythms, my work indicates that NRF1 acts as a potential master regulator of this complex system with targets in both the positive and negative feedback regulatory loops. It will be crucial to study how NRF1 interacts with CLOCK/BMAL1 heterodimers in a molecular complex and regulates the expression of other circadian regulatory genes. In addition, studies of activity patterns, sleep, and metabolism in *Nrf1* +/- heterozygous mice compared to WT controls will be of significant interest. Identification of overrepresented motifs in sequences associated with NRF1-regulated elements in circadian regulatory genes can help to identify additional potential TFs that may interact and/or function with NRF1 in the

circadian system. Other future studies would examine the role of NRF1 in SCN neurons and neuronal input and/or output pathways to the central circadian regulator.

My studies of NRF1 binding sites in HSP genes have provided substantial data for our laboratory to continue studies of the function of NRF1 in this complicated neural system. Recent studies have confirmed that another TF, SOX11, regulates *SPG4* (see **Figure 4.2**) along with NRF1 (B.J. Henson, R.D. Nicholls, unpublished data). Further, our laboratory has also identified overrepresented miRNA target sites in HSP mRNAs, predicting that the miR-96/-182 family miRNAs target nine HSP loci, an 8.5- to 11.5-fold increase over expectation (B.J. Henson, R.D. Nicholls, unpublished data). To date, our laboratory has confirmed miR-96/-182 regulation of *SPG4*, *SPG6*, *SPG1*, and of *NRF1*, with dynamic regulatory effects on mRNA and translation levels depending on cellular conditions. This work has therefore identified major TF and miRNA regulators within GRN important for neuronal function and maintenance of motor cortex axons that extend along the corticospinal tracts that degenerate in HSP. Future studies will identify additional transcriptional and miRNA regulatory elements for known HSP genes, and our laboratory plans to use the NRF1 + miR-96/-182 regulatory signature to identify and rank the top candidate genes within large chromosomal domains linked to uncloned HSP loci to aid identification of the disease genes. In the long term, this research program is identifying new functional elements for mutation studies in HSP, will provide a streamlined set of candidate genes for more rapid and cheaper identification of uncloned HSP genes, and may provide new targets for novel therapeutic approaches in neurodegenerative diseases.

Obesity and diabetes are prominent public health issues in the USA and worldwide, and are associated with cardiovascular and other diseases (Lowell & Shulman, 2005; Keller, 2006; Kouris-Blazos&Wahlqvist, 2007). It has been suggested that variation in genetic loci and

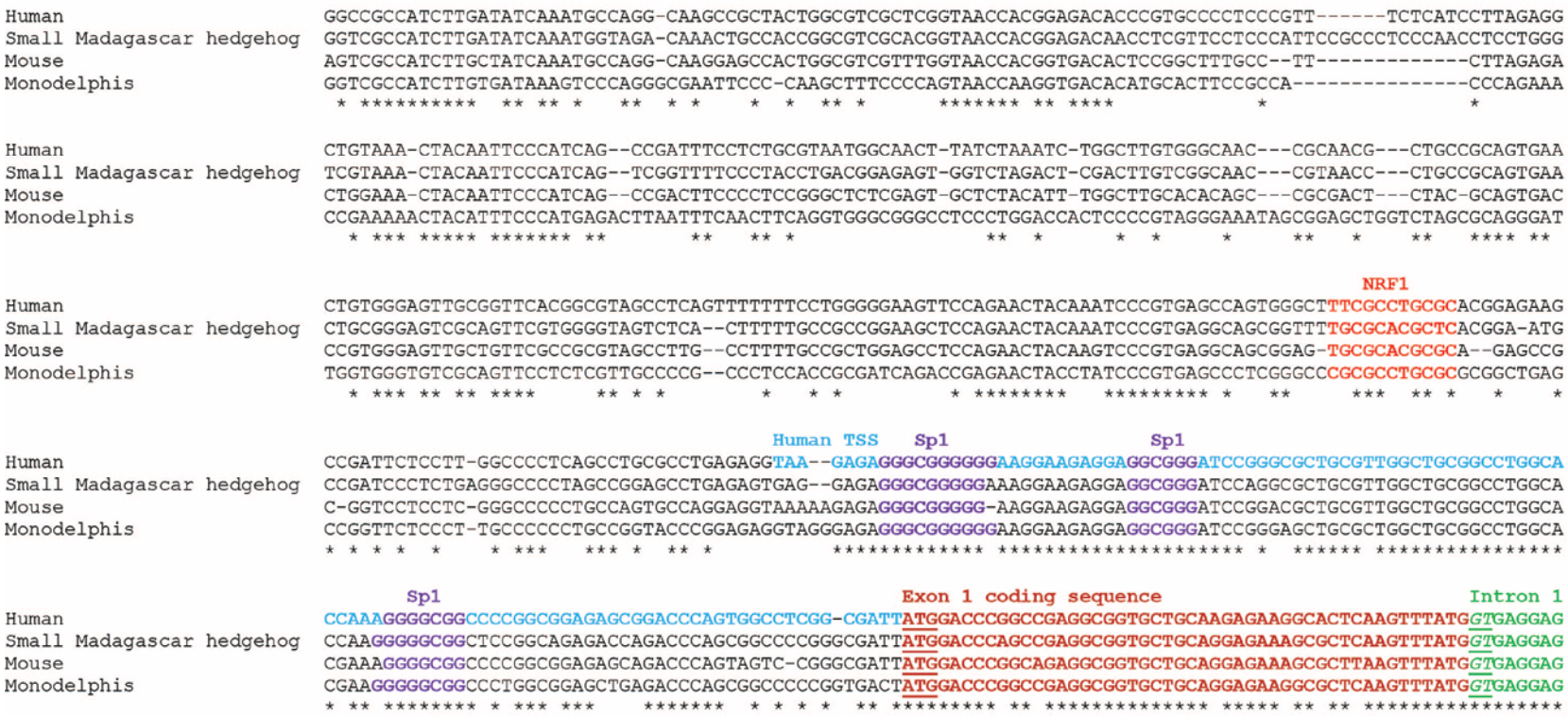
environmental factors have additive effects on metabolic and mitochondrial pathways in type II diabetes (Lowell & Shulman, 2005). Intriguingly, many genes encoding mitochondrial functions and regulators, including NRF1, have decreased expression in muscle from type II diabetes patients (Patti *et al*., 2003). A recent study has suggested genetic polymorphisms linked to *NRF1* could be a susceptibility factor for type II diabetes possibly by conferring abnormalities in triglyceride metabolism (Liu *et al*., 2008). Also, studies in our laboratory have found that NRF1 is a master regulator of gene expression in the PWS region (Stefan *et al*., in preparation), a neurobehavioral disorder associated with childhood onset of severe obesity, and that NRF1 regulates multiple genes involved in type II diabetes (M. Stefan, R.D. Nicholls, unpublished data). Furthermore, regulatory genes in circadian rhythms are also implicated in diabetes and obesity (Bass & Takahashi, 2010; Huang *et al*., 2011). Therefore, a better understanding of NRF1 target genes and pathways may provide important insights into the biological pathways and pathogenesis of diabetes, obesity, and other human disease not yet studied in detail (e.g., cancer), in addition to the roles that we have defined for NRF1 regulation in PWS, HSP, and circadian function.

APPENDIX

[MULTI-SEQUENCE ALIGNMENTS FOR 20 CIRCADIAN REGULATORY GENE PROMOTERS OR ENHANCERS]

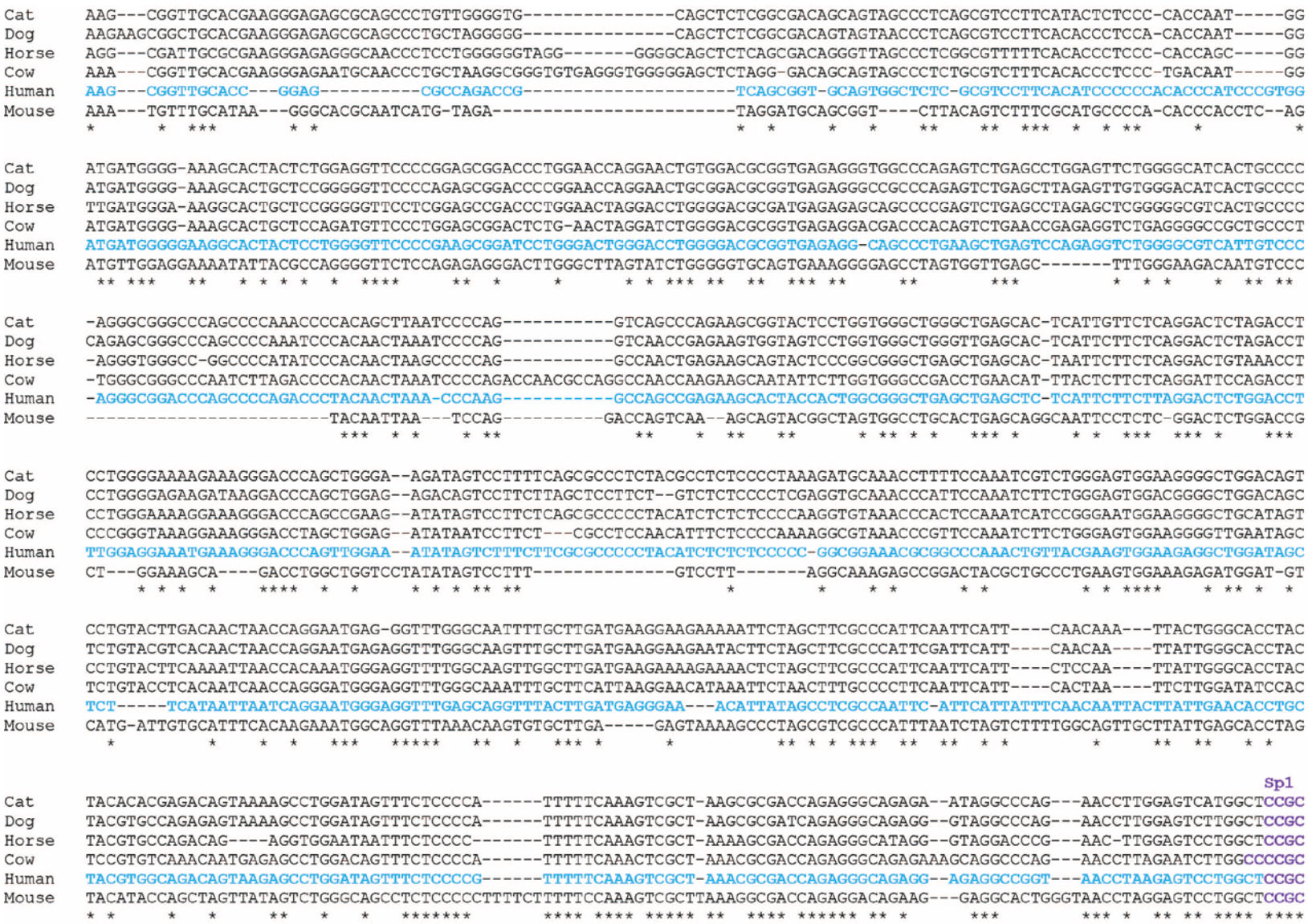
(a) *BTRC* promoter (ChIP); (b) *PPP5C* promoter (ChIP); (c) *FBXL15* promoter (ChIP); (d) *WDR5* promoter (ChIP); (e) *TIMELESS* intron 1 (ChIP); (f) *TEF* promoter (ChIP); (g) *HSF1* promoter (ChIP); (h) *EZH2* promoter (ChIP); (i) *CIPC* promoter (ChIP); (j) *CHORDC1* promoter (ChIP); (k) *ALAS1* promoter (ChIP, pub); (l) *FMR1* promoter (pub); (m) *FXR2* promoter (pub); (n) *CSNK1E* intron 1 (no exp); (o) *NR1D2* intron 1 (no exp); (p) *CREM* promoter (no exp); (q) *MYBBP1A* promoter (no exp); (r) *RAB3A* intron 1 (no exp); (s) *ATF5* intron 1 (no exp); and (t) *PROK2* promoter (no exp). Consensus sequences for CLOCK/BMAL1 (E-box), NRF1, NF-Y, RORE, DBP, and Sp1 motifs are highlighted in pink, red, green, orange, brown, and purple, respectively. ChIP indicates those *cis*-regulatory elements that we have shown bind NRF1 in SK-N-SH cells (see **Figure 3.2**). Previous publications (pub) have established that NRF1 regulates three of these promoters (for ALAS1, FMR1, and FXR2) but we present here expanded multisequence alignments. For seven genes, no experimental (no exp) data has been obtained, other than the phylogenetic conservation presented here.

(a) *BTRC* promoter (ChIP)



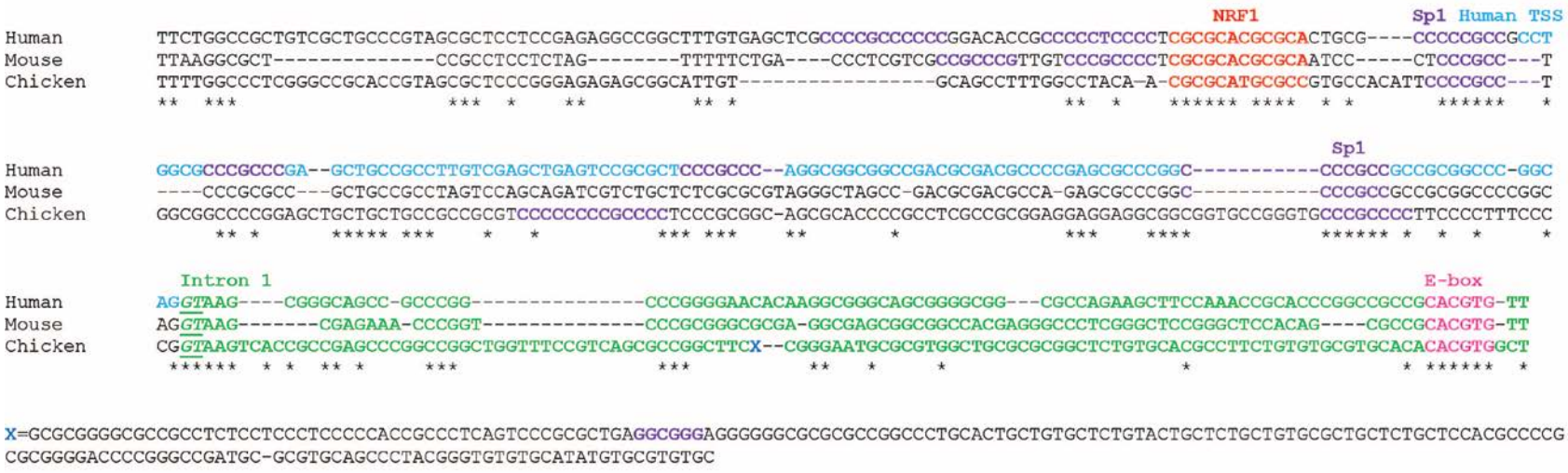
(b) *PPP5C* promoter (ChIP)

(c) *FBXL15* promoter (ChIP)


Cat
Dog
Horse
Cow
Human
Mouse
Sp1

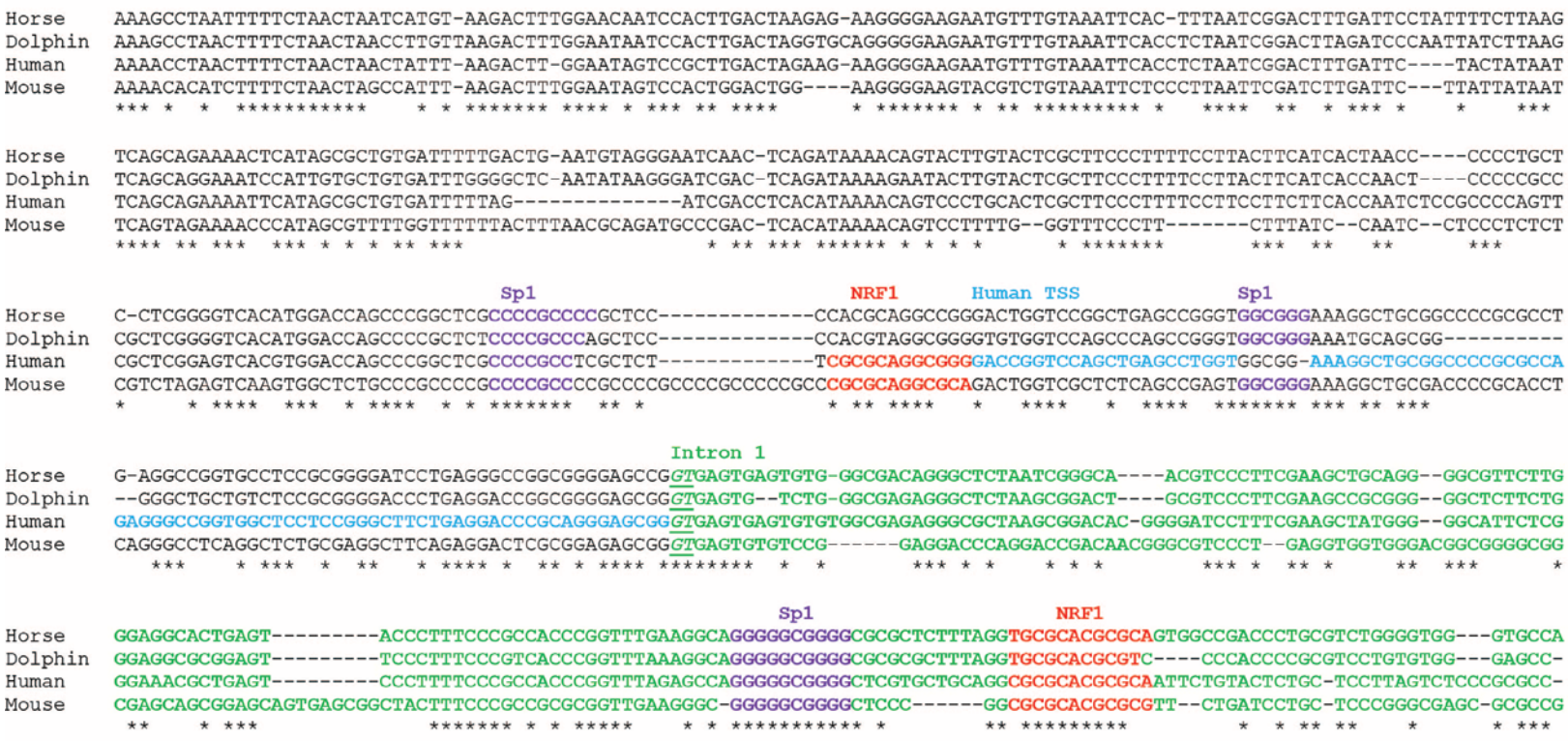

Cat
Dog
Horse
Cow
Human
Mouse
Sp1
NRF1
Exon 1 coding sequence
Intron 1
Sp-fam

(i) *WDR5* promoter (ChIP)

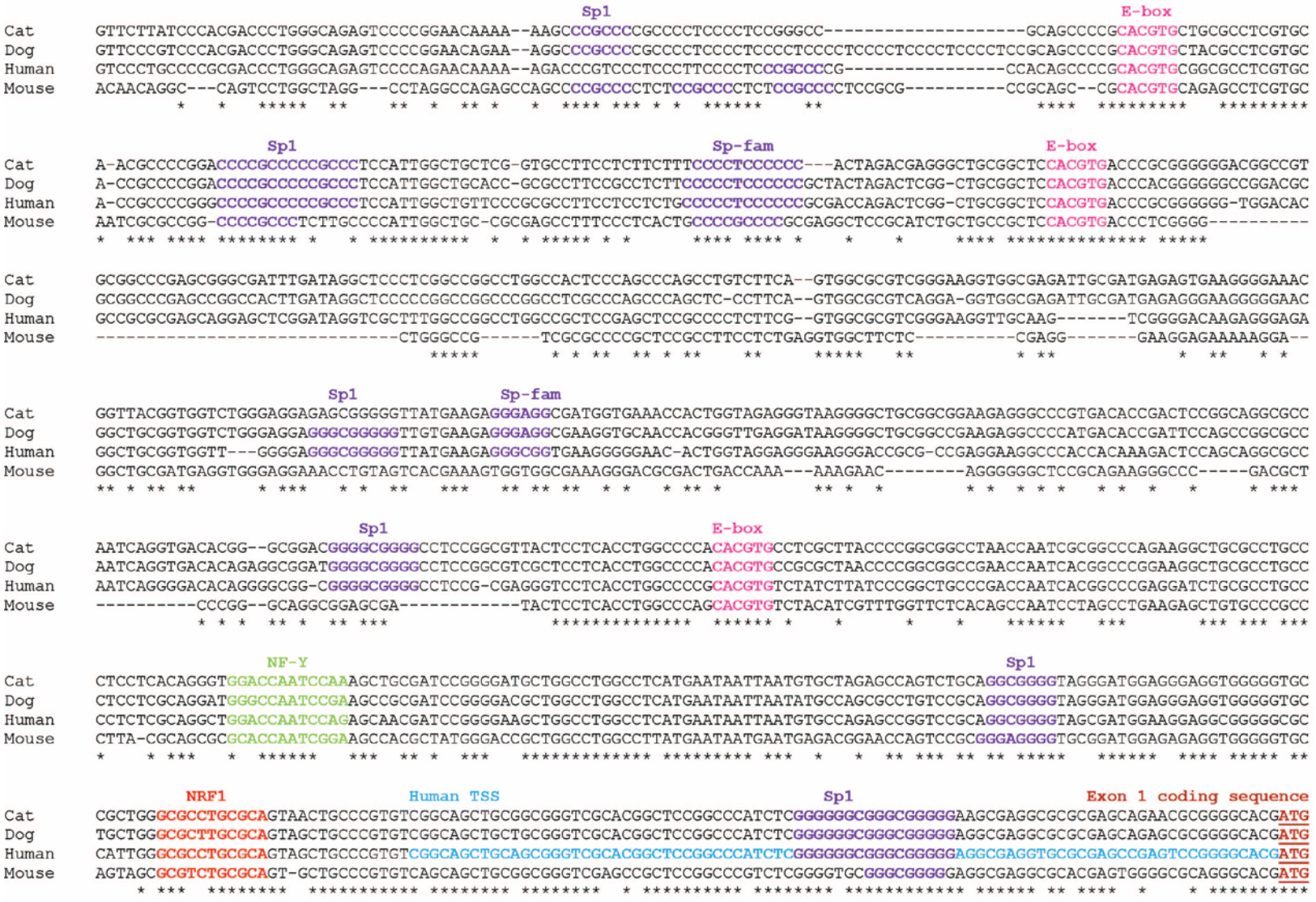


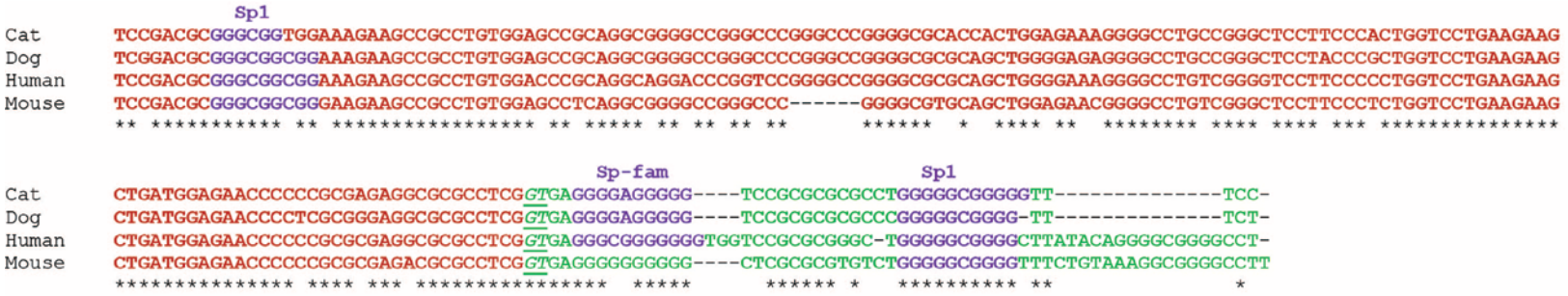
X=GCGCGGGGCGCCGCCTCTCCTCCCTCCCCCACCGCCCTCAGTCCCGCGCTGAGGCGGGAGGGGGGCGCGCGCCGGCCCTGCACTGCTGTGCTCTGTACTGCTCTGCTGTGCGCTGCTCTGCTCCACGCCCCG
CGCGGGGACCCCGGGCCGATGC-GCGTGCAGCCCTACGGGTGTGTGCATATGTGCGTGTGC

(e) *TIMELESS* intron 1 (ChIP)

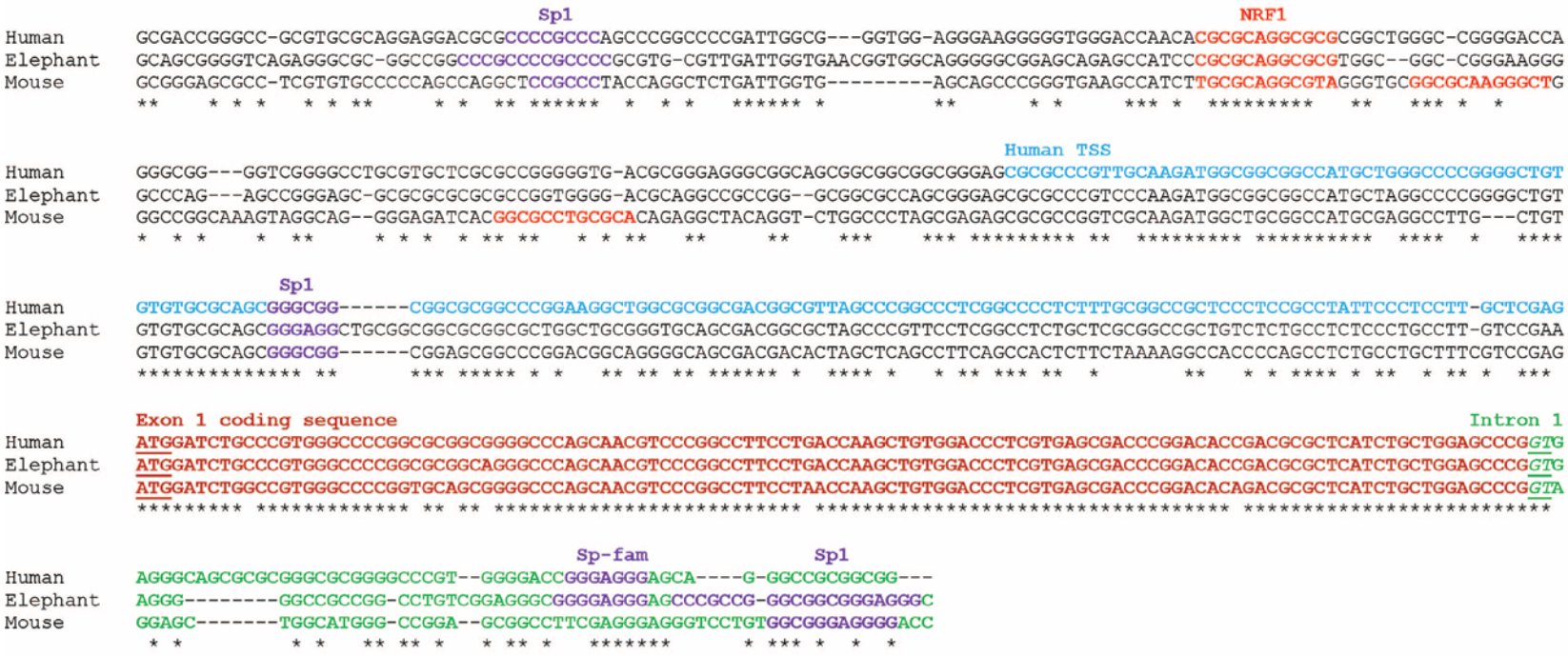


(f) *TEF* promoter (ChIP)

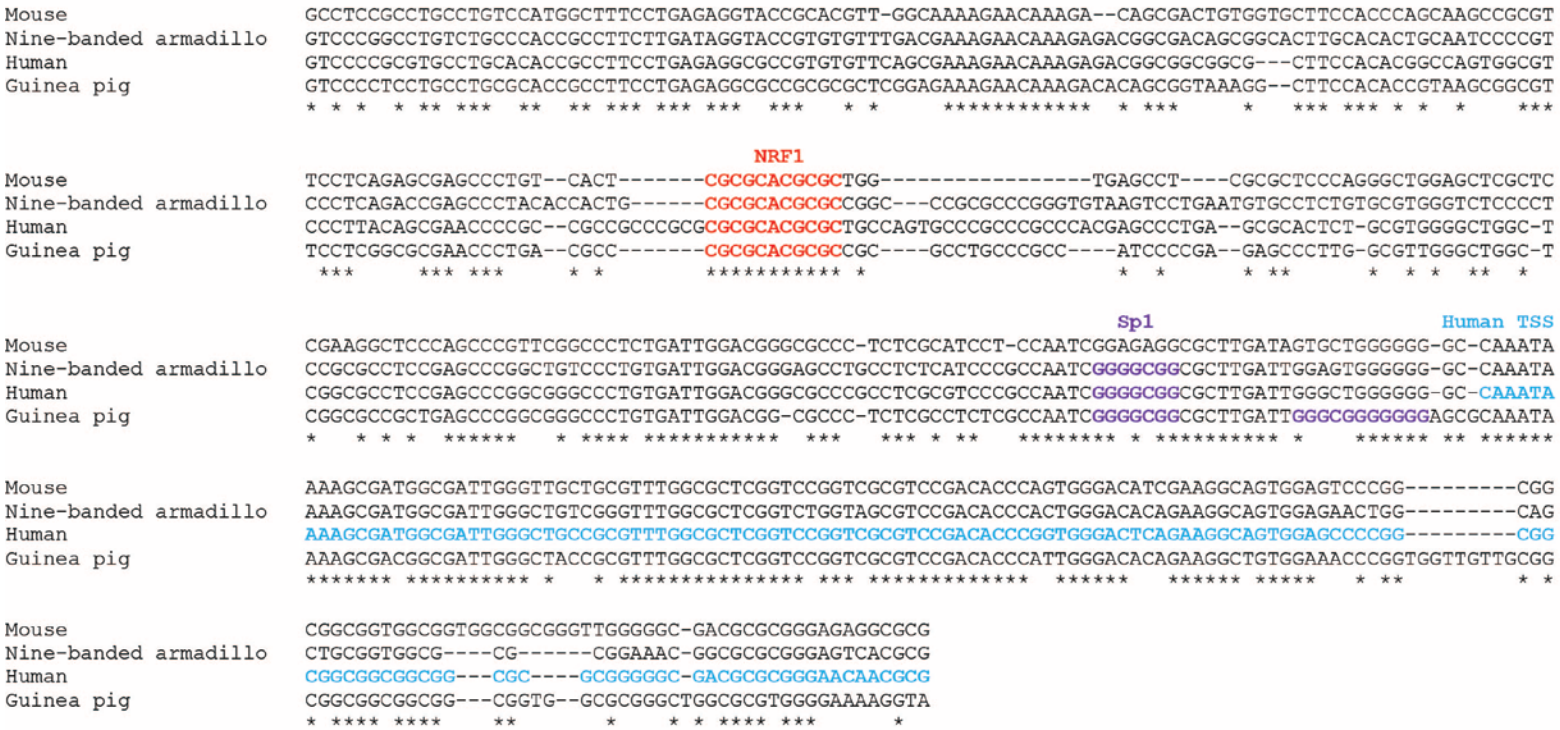




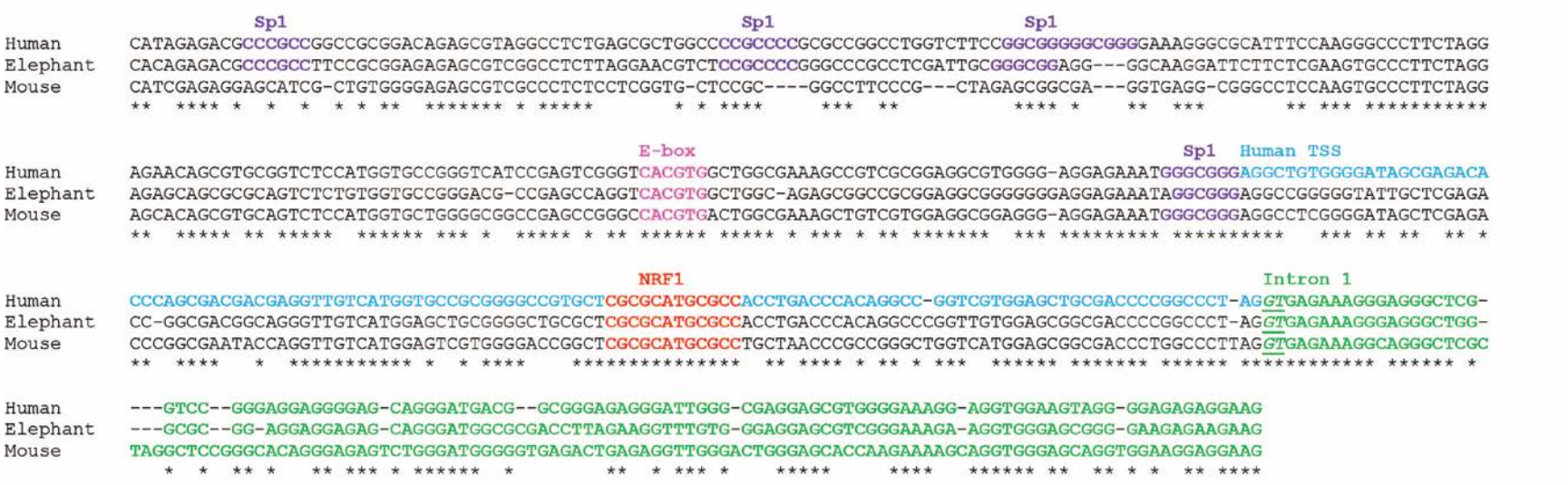
(g) *HSF1* promoter (ChIP)



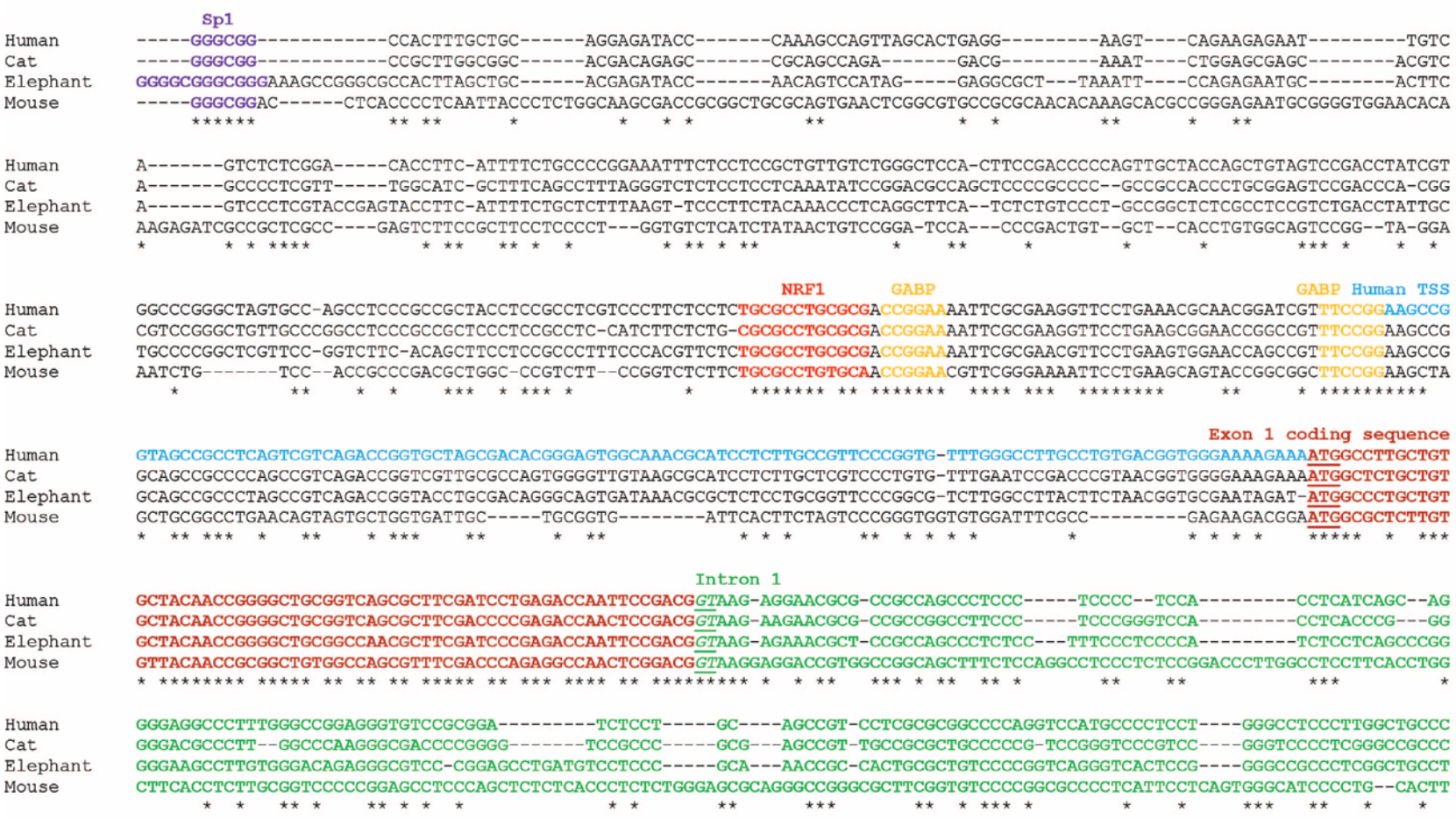
(h) *EZH2* promoter (ChIP)



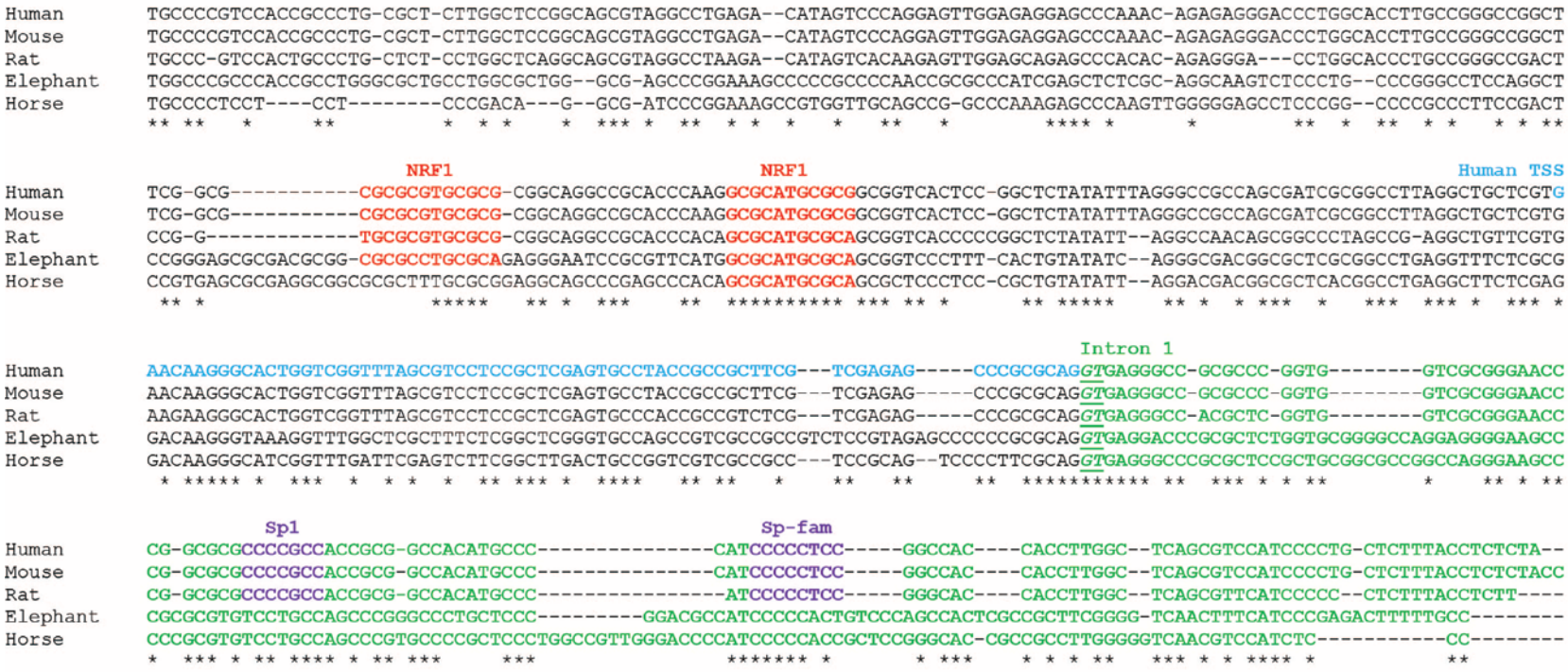
(i) *CIPC* promoter (ChIP)



(j) *CHORDC1* promoter (ChIP)

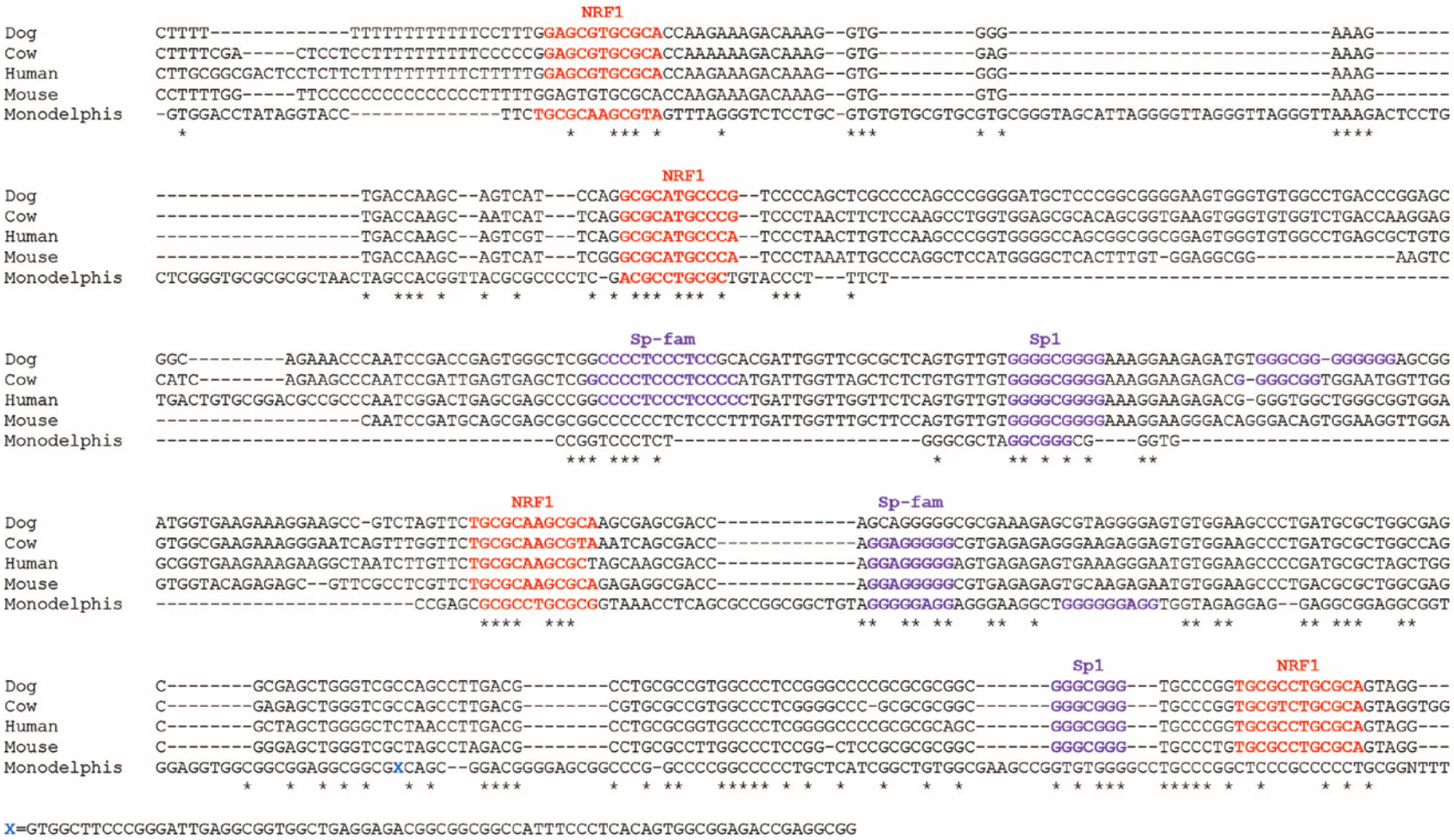


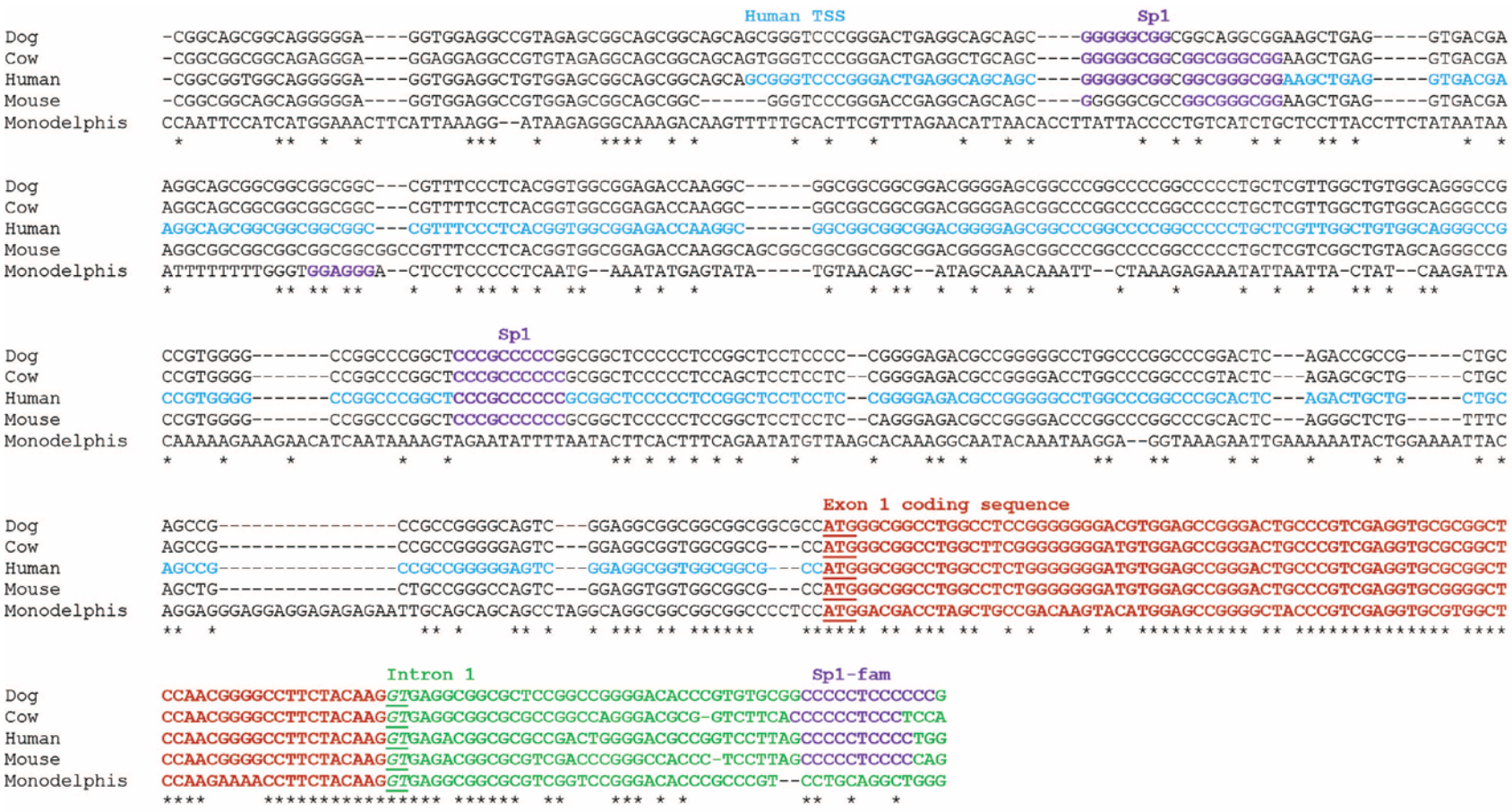
(k) *ALAS1* promoter (ChIP, pub)



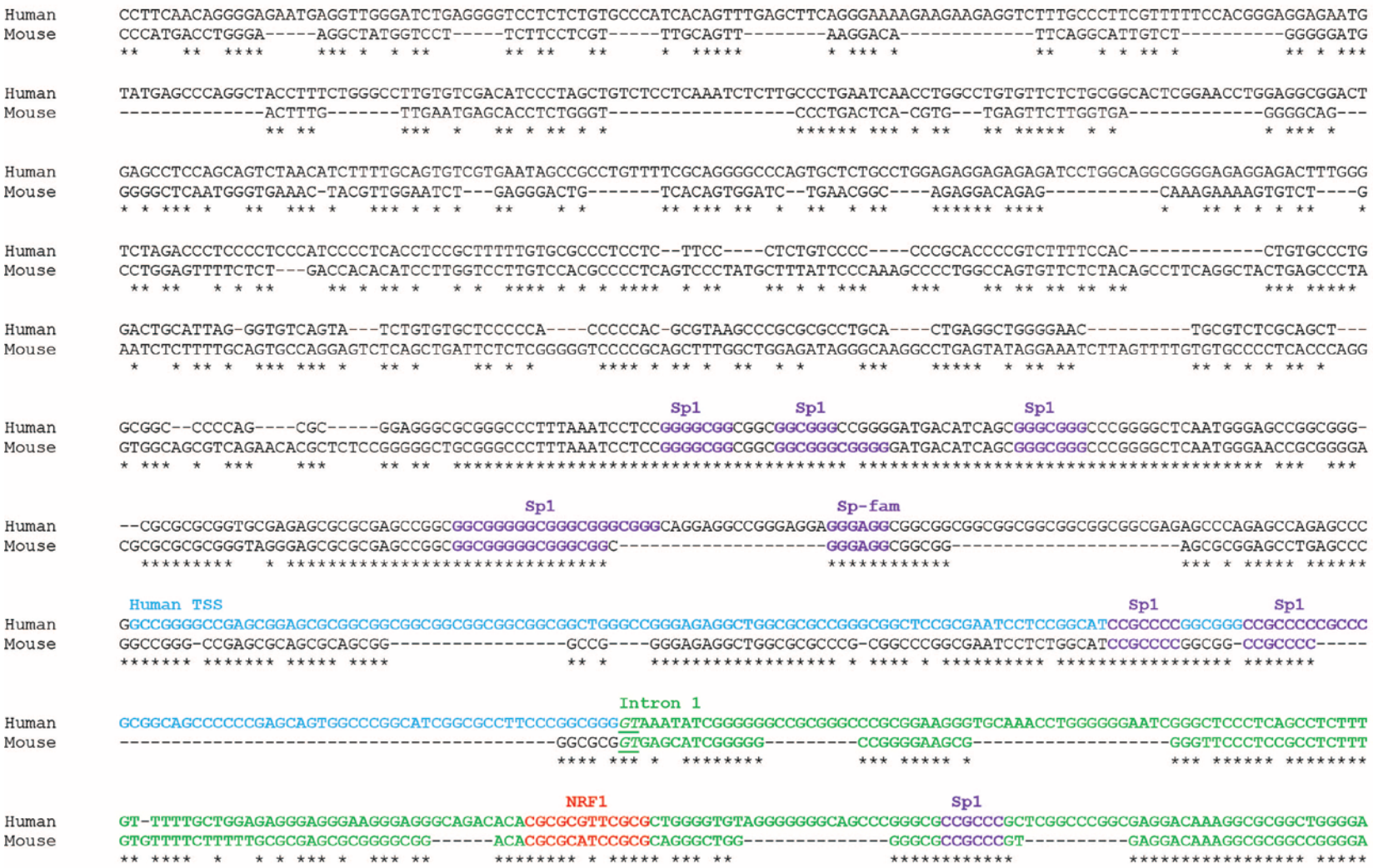
(l) *FMR1* promoter (pub)

(m) *FXR2* promoter (pub)

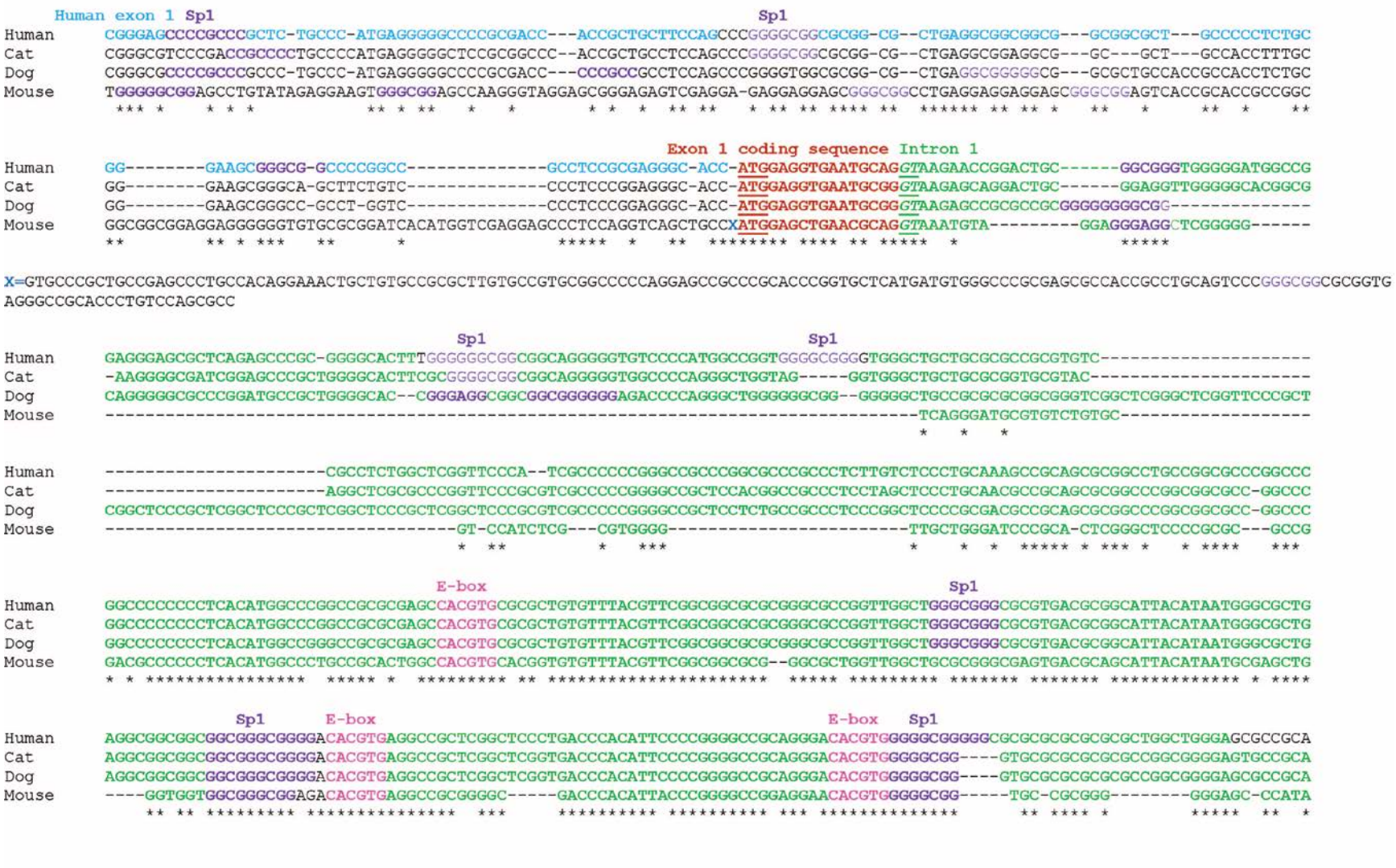


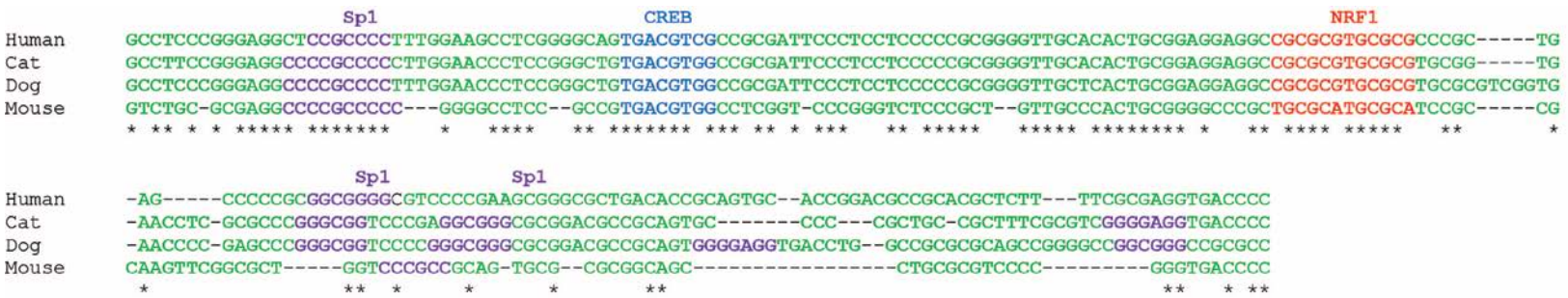

Human TSS
Sp1
Dog
Cow
Human
Mouse
Monodelphis
Sp1
Exon 1 coding sequence
Intron 1
Sp1-fam

(n) *CSNK1E* intron 1 (no exp)



(o) *NR1D2* intron 1 (no exp)



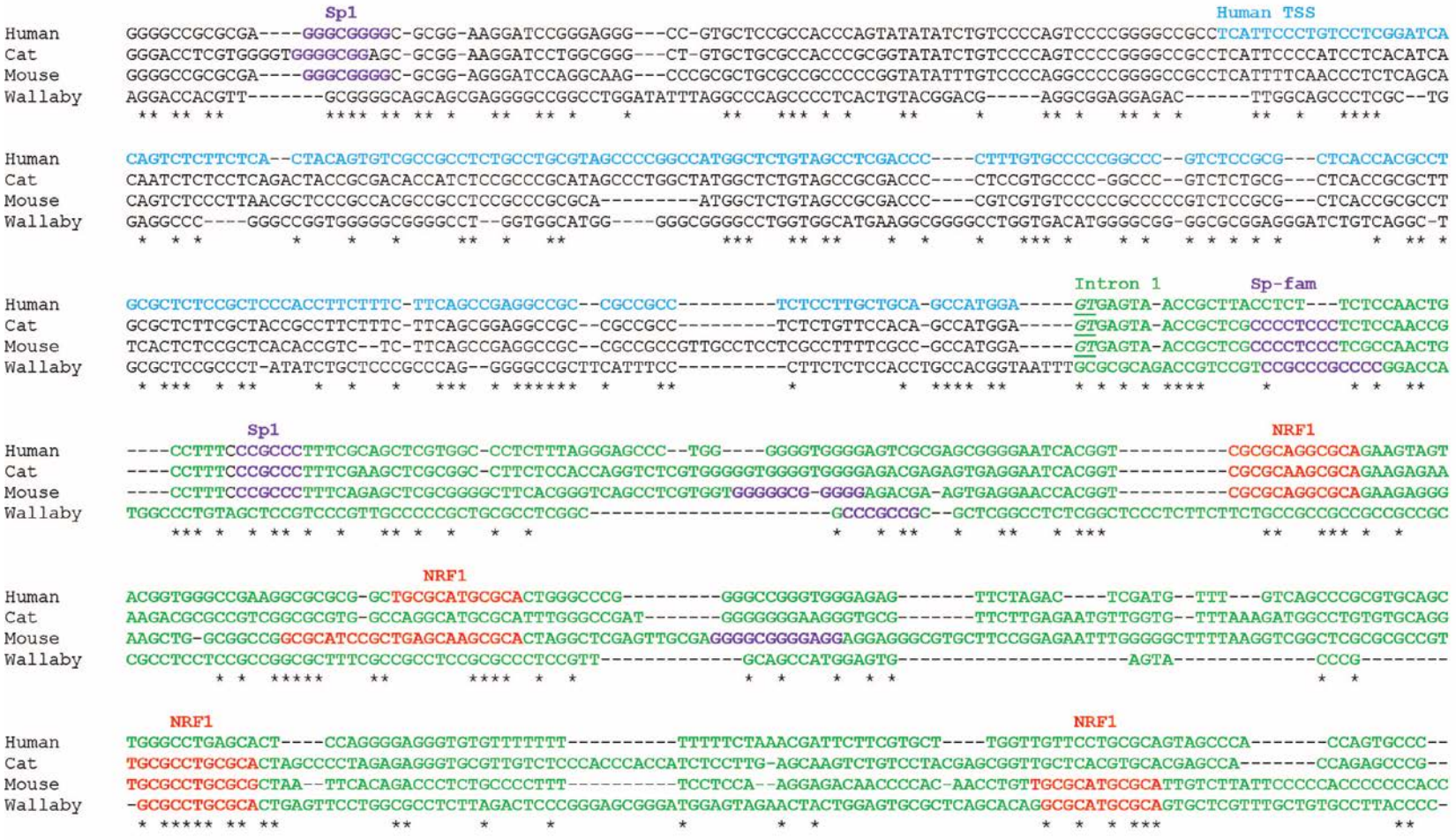


(p) *CREM* promoter (no exp)

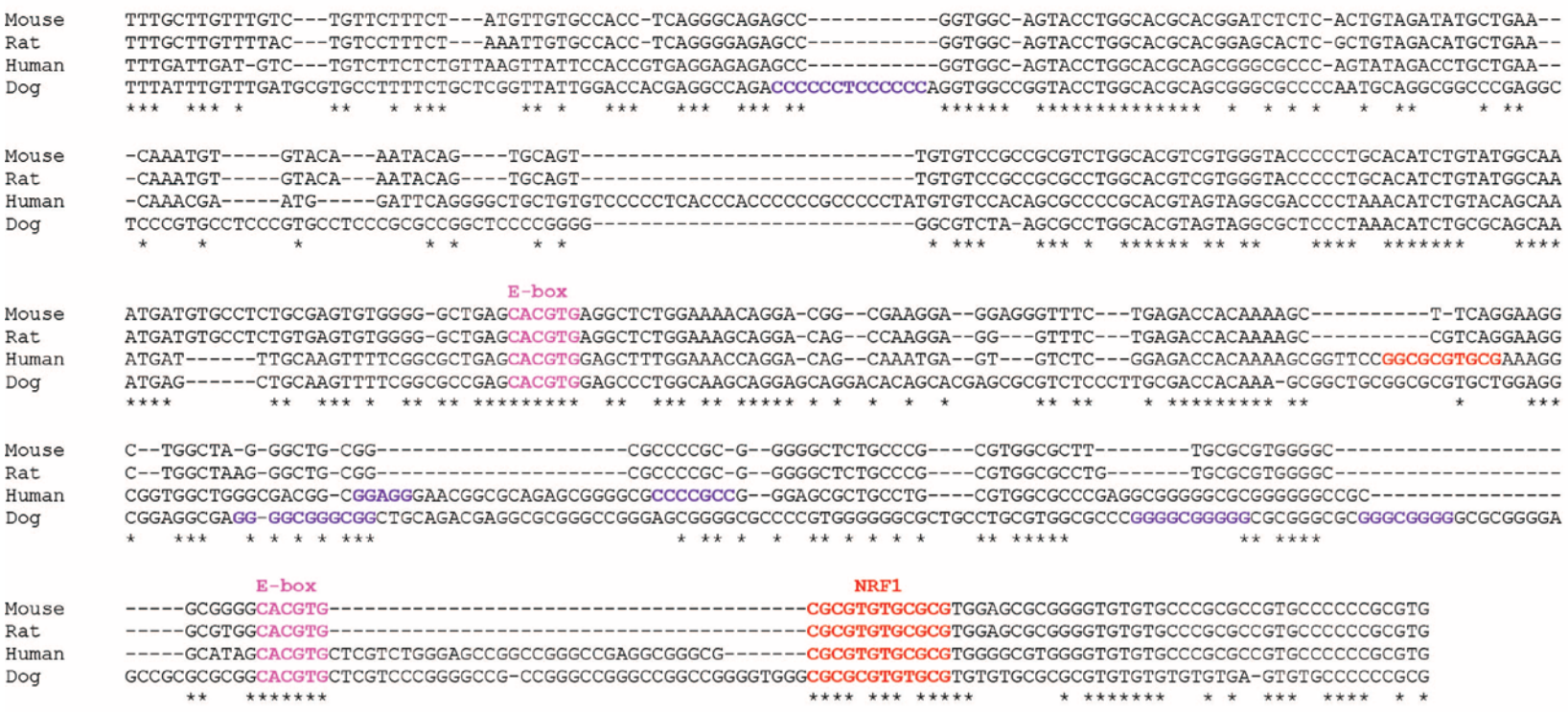
(q) *MYBBP1A* promoter (no exp)

(r) *RAB3A* intron 1 (no exp)

(s) *ATF5* intron 1 (no exp)



(t) *PROK2* promoter (no exp)



BIBLIOGRAPHY

Adelmant G, Begue A, Stehelin D, Laudet V (1996) A functional Rev-erbα responsive element located in the human Rev-erbα promoter mediates a repressing activity. *Proc Natl Acad Sci USA* 93(8):3553-3558

Akashi M, Tsuchiya Y, Yoshino T, Nishida E (2002) Control of intracellular dynamics of mammalian period proteins by casein kinase I epsilon (CKIepsilon) and CKIdelta in cultured cells. *Mol Cell Biol* 22(6):1693-1703

Akashi M, Takumi T (2005) The orphan nuclear receptor RORalpha regulates circadian transcription of the mammalian core-clock Bmal1. *Nat Struct Mol Biol* 12(5):441-448

Akten B, Jauch E, Genova GK, Kim EY, Edery I, Raabe T, Jackson FR (2003) A role for CK2 in the Drosophila circadian oscillator. *Nat Neurosci* 6(3):251-257

Amelio AL, Miraglia LJ, Conkright JJ, Mercer BA, Batalov S, Cavett V, Orth AP, Busby J, Hogenesch JB, Conkright MD (2007) A coactivator trap identifies NONO (p54nrb) as a component of the cAMP-signaling pathway. *Proc Natl Acad Sci USA* 104(51):20314-20319

Amor DJ, Halliday J (2008) A review of known imprinting syndromes and their association with assisted reproduction technologies. *Hum Reprod* 23(12):2826-2834

Andersson U, Scarpulla RC (2001) Pgc-1-related coactivator, a novel, serum-inducible coactivator of nuclear respiratory factor 1-dependent transcription in mammalian cells. *Mol Cell Biol* 21(11):3738-3749

Asher G, Gatfield D, Stratmann M, Reinke H, Dibner C, Kreppel F, Mostoslavsky R, Alt FW, Schibler U (2008) SIRT1 regulates circadian clock gene expression through PER2 deacetylation. *Cell* 134(2):317-328

Aviram R, Kidron D, Silverstein S, Lerer I, Abeliovich D, Tepper R, Dolfin Z, Markovitch O, Arnon S (2008) Placental mesenchymal dysplasia associated with transient neonatal diabetes mellitus and paternal UPD6. *Placenta* 29(7):646-649

Baar K, Wende AR, Jones TE, Marison M, Nolte LA, Chen M, Kelly DP, Holloszy JO (2002) Adaptations of skeletal muscle to exercise: rapid increase in the transcriptional coactivator PGC-1. *FASEB J* 16(14):1879-1886

Balsalobre A, Damiola F, Schibler U (1998) A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* 93(6):929-937

Balsalobre A, Brown SA, Marcacci L, Tronche F, Kellendonk C, Reichardt HM, Schütz G, Schibler U (2000) Resetting of circadian time in peripheral tissues by glucocorticoid signaling. *Science* 289(5488):2344-2347

Ban HJ, Kim SC, Seo J, Kang HB, Choi JK (2011) Genetic and metabolic characterization of insomnia. *PLoS One* 6(4):e18455

Barlow DP, Stöger R, Herrmann BG, Saito K, Schweifer N (1991) The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature* 349(6304):84-87

Bass J, Takahashi JS (2010) Circadian integration of metabolism and energetics. *Science* 330(6009):1349-1354

Becker TS, Burgess SM, Amsterdam AH, Allende ML, Hopkins N (1998) not really finished is crucial for development of the zebrafish outer retina and encodes a transcription factor highly homologous to human Nuclear Respiratory Factor-1 and avian Initiation Binding Repressor. *Development* 125(22):4369-4378

Bergeron R, Ren JM, Cadman KS, Moore IK, Perret P, Pypaert M, Young LH, Semenkovich CF, Shulman GI (2001) Chronic activation of AMP kinase results in NRF-1 activation and mitochondrial biogenesis. *Am J Physiol-Endocrin Metab* 281(6):E1340-E1346

Blackstone C, O'Kane CJ, Reid E (2011) Hereditary spastic paraplegias: membrane traffic and the motor pathway. *Nat Rev Neurosci* 12(1):31-42

Blau J (2008) PERspective on PER phosphorylation. *Genes Dev* 22(13):1737-1740

Boccaccio I, Glatt-Deeley H, Watrin F, Roeckel N, Lalande M, Muscatelli F (1999) The human MAGEL2 gene and its mouse homologue are paternally expressed and mapped to the Prader-Willi region. *Hum Mol Genet* 8(13):2497-2505

Boden G, Chen X, Polansky M (1999) Disruption of circadian insulin secretion is associated with reduced glucose uptake in first-degree relatives of patients with type 2 diabetes. *Diabetes* 48(11):2182-2188

Bogarad LD, Arnone MI, Chang C, Davidson EH (1998) Interference with gene regulation in living sea urchin embryos: transcription factor knock out (TKO), a genetically controlled

vector for blockade of specific transcription factors. *Proc Natl Acad Sci U S A* 95(25):14827-14832

Bozek K, Rosahl AL, Gaub S, Lorenzen S, Herzel H (2010) Circadian transcription in liver. *Biosystems* 102(1):61-69

Brown SA, Ripperger J, Kadener S, Fleury-Olela F, Vilbois F, Rosbash M, Schibler U (2005) PERIOD1-associated proteins modulate the negative limb of the mammalian circadian oscillator. *Science* 308(5722):693-696

Brummelkamp TR, Bernards R, Agami R (2002) A system for stable expression of short interfering RNAs in mammalian cells. *Science* 296(5567):550-553

Buiting K, Nazlican H, Galetzka D, Wawrzik M, Gross S, Horsthemke B (2007) C15orf2 and a novel non-coding transcript from the Prader-Willi/Angelman syndrome region show monoallelic expression in fetal brain. *Genomics* 89(5):588-595

Buiting K (2010) Prader-Willi syndrome and Angelman syndrome. *Am J Med Genet Part C Semin Med Genet* 154C(3):365-376

Busino L, Bassermann F, Maiolica A, Lee C, Nolan PM, Godinho SI, Draetta GF, Pagano M (2007) SCFFbxl3 controls the oscillation of the circadian clock by directing the degradation of cryptochrome proteins. *Science* 316(5826):900-904

Butler JV, Whittington JE, Holland AJ, Boer H, Clarke D, Webb T (2002) Prevalence of, and risk factors for, physical ill-health in people with Prader-Willi syndrome: a population-based study. *Dev Med Child Neurol* 44(4):248-255

Butler MP, Honma S, Fukumoto T, Kawamoto T, Fujimoto K, Noshiro M, Kato Y, Honma K (2004) Dec1 and Dec2 expression is disrupted in the suprachiasmatic nuclei of Clock mutant mice. *J Biol Rhythms* 19(2):126-134

Calzone FJ, Höög C, Teplow DB, Cutting AE, Zeller RW, Britten RJ, Davidson EH (1991) Gene regulatory factors of the sea urchin embryo. I. Purification by affinity chromatography and cloning of P3A2, a novel DNA-binding protein. *Development* 112(1):335-350

Cam H, Balciunaite E, Blais A, Spektor A, Scarpulla RC, Young R, Kluger Y, Dynlacht BD (2004) A common set of gene regulatory networks links metabolism and growth inhibition. *Mol Cell* 16(3):399-411

Capecchi MR (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* 6(6):507-512

Cardone L, Hirayama J, Giordano F, Tamaru T, Palvimo JJ, Sassone-Corsi P (2005) Circadian clock control by SUMOylation of BMAL1. *Science* 309(5739):1390-1394

Cavaillé J, Nicoloso M, Bachellerie JP (1996) Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature* 383(6602):732-735

Cavaillé J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Hüttenhofer A (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci USA* 97(26):14311-14316

Chai JH, Locke DP, Ohta T, Greally JM, Nicholls RD (2001) Retrotransposed genes such as Frat3 in the mouse Chromosome 7C Prader-Willi syndrome region acquire the imprinted status of their insertion site. *Mamm Genome* 12(11): 813-821

Chang WT, Huang AM (2004) α-Pal/NRF-1 regulates the promoterof the human integrin-associated protein/CD47 gene. *J Biol Chem* 279(15):14542-14550

Chang WT, Chen HI, Chiou RJ, Chen CY, Huang AM (2005) A novel function of transcription factor alpha-Pal/NRF-1: increasing neurite outgrowth. *Biochem Biophys Res Commun* 334(1):199-206

Chappell PE, White RS, Mellon PL (2003) Circadian gene expression regulates pulsatile gonadotropin-releasing hormone (GnRH) secretory patterns in the hypothalamic GnRH-secreting GT1-7 cell line. *J Neurosci* 23(35):11202-11213

Chen S, Nagy PL, Zalkin H (1997) Role of NRF-1 in bidirectional transcription of the human GPAT-AIRC purine biosynthesis locus. *Nucleic Acids Res* 25(9):1809-1816

Chen K, Rajewsky N (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 8(2):93-103

Cheng MY, Bullock CM, Li C, Lee AG, Bermak JC, Belluzzi J, Weaver DR, Leslie FM, Zhou QY (2002) Prokineticin 2 transmits the behavioural circadian rhythm of the suprachiasmatic nucleus. *Nature* 417(6887):405-410

Cheng MY, Bittman EL, Hattar S, Zhou QY (2005) Regulation of prokineticin 2 expression by light and the circadian clock. *BMC Neurosci* 6:17

Cheung TH, Barthel KK, Kwan YL, Liu X (2007) Identifying pattern-defined regulatory islands in mammalian genomes. *Proc Natl Acad Sci USA* 104(24):10116-10121

Chi SW, Zang JB, Mele A, Darnell RB (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 460(7254):479-486

Chudoba I, Franke Y, Senger G, Sauerbrei G, Demuth S, Beensen V, Neumann A, Hansmann I, Claussen U (1999) Maternal UPD 20 in a hyperactive child with severe growth retardation. *Eur J Hum Genet* 7(5):533-540

Ciccone DN, Chen T (2009) Histone lysine methylation in genomic imprinting. *Epigenetics* 4(4):216-220

Coon SL, Weller JL, Korf HW, Namboodiri MA, Rollag M, Klein DC (2001) cAmp regulation of arylalkylamine N-acetyltransferase (AANAT, EC 2.3.1.87): a new cell line (1E7) provides evidence of intracellular AANAT activation. *J Biol Chem* 276(26):24097-24107

Cos RP, Drauss MR, Balis ME, Dancis J (1974) Letter: Mouse fibroblasts A9 are deficient in HPRT and APRT. *Am J Hum Genet* 26(2):272-273

Crosio C, Cermakian N, Allis CD, Sassone-Corsi P (2000) Light induces chromatin modification in cells of the mammalian circadian clock. *Nat Neurosci* 3(12):1241-1247

Cunningham JT, Rodgers JT, Arlow DH, Vazquez F, Mootha VK, Puigserver P (2007) mTOR controls mitochondrial oxidative function through a YY1 – PGC-1 α transcriptional complex. *Nature* 450(7170):736-740

Curtis AM, Seo SB, Westgate EJ, Rudic RD, Smyth EM, Chakravarti D, FitzGerald GA, McNamara P (2004) Histone acetyltransferase-dependent chromatin remodeling and the vascular clock. *J Biol Chem* 279(8):7091-7097

Darlington TK, Wager-Smith K, Ceriani MF, Staknis D, Gekakis N, Steeves TD, Weitz CJ, Takahashi JS, Kay SA (1998) Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim. *Science* 280(5369):1599-1603

DeChiara TM, Robertson EJ, Efstratiadis A (1991) Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* 64(4):849-859

de los Santos T, Schweizer J, Rees CA, Francke U (2000) Small evolutionarily conserved RNA, resembling C/D box small nucleolar RNA, is transcribed from PWCR1, a novel imprinted gene in the Prader-Willi deletion region, which Is highly expressed in brain. *Am J Hum Genet* 67(5):1067-1082

Depienne C, Stevanin G, Brice A, Durr A (2007) Hereditary spastic paraplegias: an update. *Curr Opin Neurol* 20(6):674-680

DeSimone SM, White K (1993) The *Drosophila* erect wing gene, which is important for both neuronal and muscle development, encodes a protein which is similar to the sea urchin P3A2 DNA binding protein. *Mol Cell Biol* 13(6):3641-3649

Dhar SS, Ongwijitwat S, Wong-Riley MT (2008) Nuclear respiratory factor 1 regulates all ten nuclear-encoded subunits of cytochrome c oxidase in neurons. *J Biol Chem* 283(6):3120-3129

Dhar SS, Ongwijitwat S, Wong-Riley MT (2009) Chromosome conformation capture of all 13 genomic Loci in the transcriptional regulation of the multisubunit bigenomic cytochrome C oxidase in neurons. *J Biol Chem* 284(28):18644-18650

Dhar SS, Wong-Riley MT (2009) Coupling of energy metabolism and synaptic transmission at the transcriptional level: role of nuclear respiratory factor 1 in regulating both cytochrome c oxidase and NMDA glutamate receptor subunit genes. *J Neurosci* 29(2):483-492

Ding F, Li HH, Zhang S, Solomon NM, Camper SA, Cohen P, Francke U (2008) SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice. *PLoS ONE* 3(3):e1709

Dion PA, Daoud H, Rouleau GA (2009) Genetics of motor neuron disorders: new insights into pathogenic mechanisms. *Nat Rev Genet* 10(11):769-782

Dockendorff TC, Su HS, McBride SM, Yang Z, Choi CH, Siwicki KK, Sehgal A, Jongens TA (2002) Drosophila lacking dfmr1 activity show defects in circadian output and fail to maintain courtship interest. *Neuron* 34(6):973-984

Doi M, Okano T, Yujnovsky I, Sassone-Corsi P, Fukada Y (2004) Negative control of circadian clock regulator E4BP4 by casein kinase Iepsilon-mediated phosphorylation. *Curr Biol* 14(11):975-980

Doi M, Hirayama J, Sassone-Corsi P (2006) Circadian regulator CLOCK is a histone acetyltransferase. *Cell* 125(3):497-508

Dubchak I, Brudno M, Loots GG, Pachter L, Mayor C, Rubin EM, Frazer KA (2000) Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res* 10(9):1304-1306

Earnest DJ, Cassone VM (2005) Cell culture models for oscillator and pacemaker function: recipes for dishes with circadian clocks? *Methods Enzymol* 393:558-578

Efiok BJ, Chiorini JA, Safer B (1994) A key transcription factor for eukaryotic initiation factor-2 alpha is strongly homologous to developmental transcription factors and may link metabolic genes to cellular growth and development. *J Biol Chem* 269(29):18921-18930

Efiok BJ, Safer B (2000) Transcriptional regulation of E2F-1 and eIF-2 genes by alpha-pal: a potential mechanism for coordinated regulation of protein synthesis, growth, and the cell cycle. *Biochim Biophys Acta* 1495(1):51-68

Eggermann T, Mergenthaler S, Eggermann K, Albers A, Linnemann K, Fusch C, Ranke MB, Wollmann HA (2001) Identification of interstitial maternal uniparental disomy (UPD) (14) and complete maternal UPD(20) in a cohort of growth retarded patients. *J Med Genet* 38(2):86-89

Eggermann T (2010) Russell-Silver syndrome. *Am J Med Genet C Semin Med Genet* 154C(3):355-364

Eide EJ, Vielhaber EL, Hinz WA, Virshup DM (2002) The circadian regulatory proteins BMAL1 and cryptochromes are substrates of casein kinase Iepsilon. *J Biol Chem* 77(19):17248-17254

Etchegaray JP, Lee C, Wade PA, Reppert SM (2003) Rhythmic histone acetylation underlies transcription in the mammalian circadian clock. *Nature* 421(6919):177-182

Etchegaray JP, Yang X, DeBruyne JP, Peters AH, Weaver DR, Jenuwein T, Reppert SM (2006) The polycomb group protein EZH2 is required for mammalian circadian clock function. *J Biol Chem* 281(30):21209-21215

Etchegaray JP, Machida KK, Noton E, Constance CM, Dallmann R, Di Napoli MN, DeBruyne JP, Lambert CM, Yu EA, Reppert SM, Weaver DR (2009) Casein kinase 1 delta regulates the pace of the mammalian circadian clock. *Mol Cell Biol* 29(14):3853-3866

Evans MJ, Scarpulla RC (1989) Interaction of nuclear factors with multiple sites in the somatic cytochrome c promoter. Characterization of upstream NRF-1, ATF, and intron Sp1 recognition sequences. *J Biol Chem* 264(24):14361-14368

Evans MJ, Scarpulla RC (1990) NRF-1: a trans-activator of nuclear-encoded respiratory genes in animal cells. *Genes Dev* 4(6):1023-1034

Färber C, Dittrich B, Buiting K, Horsthemke B (1999) The chromosome 15 imprinting centre (IC) region has undergone multiple duplication events and contains an upstream exon of SNRPN that is deleted in all Angelman syndrome patients with an IC microdeletion. *Hum Mol Genet* 8(2):337-343

Färber C, Gross S, Neesen J, Buiting K, Horsthemke B (2000) Identification of a testis-specific gene (C15orf2) in the Prader-Willi syndrome region on chromosome 15. *Genomics* 65(2):174-183

Fazio IK, Bolger TA, Gill G (2001) Conserved regions of the Drosophila erect wing protein contribute both positively and negatively to transcriptional activity. *J Biol Chem* 276(22):18710-18716

Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. *Nature* 447(7143):433-440

Fernández-Rebollo E, Lecumberri B, Garin I, Arroyo J, Bernal-Chico A, Goñi F, Orduña R, Spanish PHP Group, Castaño L, de Nanclares GP (2010) New mechanisms involved in paternal 20q disomy associated with pseudohypoparathyroidism. *Eur J Endocrinol* 163(6):953-962

Fink JK (2009) Hereditary Spastic Paraplegia Overview. In: Pagon RA, Bird TC, Dolan CR, Stephens K, editors. *GeneReviews [Internet]*. Seattle (WA): University of Washington, Seattle; 1993-2000 Aug 15 [updated 2009 Feb 3]

FitzGerald PC, Shlyakhtenko A, Mir AA, Vinson C (2004) Clustering of DNA sequences in human promoters. *Genome Res* 14(8):1562-1574

Fleming RJ, DeSimone SM, White K (1989) Molecular isolation and analysis of the erect wing locus in Drosophila melanogaster. *Mol Cell Biol* 9(2):719-725

Fonjallaz P, Ossipow V, Wanner G, Schibler U (1996) The two PAR leucine zipper proteins, TEF and DBP, display similar circadian and tissue-specific expression, but have different target promoter preferences. *EMBO J* 15(2):351-362

Foulkes NS, Duval G, Sassone-Corsi P (2006) Adaptive inducibility of CREM as transcriptional memory of circadian rhythms. *Nature* 381(6577):83-85

Fried M, Crothers DM (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res* 9(23):6505-6525

Fu L, Pelicano H, Liu J, Huang P, Lee C (2002) The circadian gene Period2 plays an important role in tumor suppression and DNA damage response in vivo. *Cell* 111(1):41-50

Gabriel JM, Higgins MJ, Gebuhr TC, Shows TB, Saitoh S, Nicholls RD (1998) A model system to study genomic imprinting of human genes. *Proc Natl Acad Sci U S A* 95(25):14857-14862

Gallego M, Virshup DM. (2007) Post-translational modifications regulate the ticking of the circadian clock. *Nat Rev Mol Cell Biol* 8(2):139-148

Garbarino-Pico E, Niu S, Rollag MD, Strayer CA, Besharse JC, Green CB (2007) Immediate early response of the circadian polyA ribonuclease nocturnin to two extracellular stimuli. *RNA* 13(5):745-755

Garcia-Roves PM, Osler ME, Holmstrom MH, Zierath JR (2008) Gain-of-function R225Q Mutation in AMPactivated protein kinase gamma 3 subunit increases mitochondrial biogenesis in glycolytic skeletal muscle. *J Biol Chem* 283(51):35724-35734

Gaulton KJ, Willer CJ, Li Y, Scott LJ, Conneely KN, Jackson AU, Duren WL, Chines PS, Narisu N, Bonnycastle LL, Luo J, Tong M, Sprau AG, Pugh EW, Doheny KF, Valle TT, Abecasis GR, Tuomilehto J, Bergman RN, Collins FS, Boehnke M, Mohlke KL (2008) Comprehensive association study of type 2 diabetes and related quantitative traits with 222 candidate genes. *Diabetes* 57(11):3136-3144

Gekakis N, Staknis D, Nguyen HB, Davis FC, Wilsbacher LD, King DP, Takahashi JS, Weitz CJ (1998) Role of the CLOCK protein in the mammalian circadian mechanism. *Science* 280(5369):1564-1569

Gerstner JR, Landry CF (2007) The zinc-binding protein chordc1 undergoes complex diurnal changes in mRNA expression during mouse brain development. *Neurochem Res* 32(2):241-250

Gómez-Cuadrado A, Martín M, Noël M, Ruiz-Carrillo A (1995) Initiation binding repressor, a factor that binds to the transcription initiation site of the histone h5 gene, is a glycosylated member of a family of cell growth regulators [corrected]. *Mol Cell Biol* 15(12):6670-6685

Gopalakrishnan L, Scarpulla RC (1994) Differential regulation of respiratory chain subunits by a CREB-dependent signal transduction pathway. Role of cyclic AMP in cytochrome c and COXIV gene expression. *J Biol Chem* 269(1):105-113

Gopalakrishnan L, Scarpulla RC (1995) Structure, expression, and chromosomal assignment of the human gene encoding nuclear respiratory factor 1. *J Biol Chem* 270(30):18019-18025

Goutsias J, Lee NH (2007) Computational and experimental approaches for modeling gene regulatory networks. *Curr Pharm Des* 13(14):1415-1436

Gravotta L, Gavrila AM, Hood S, Amir S (2011) Global Depletion of Dopamine Using Intracerebroventricular 6-Hydroxydopamine Injection Disrupts Normal Circadian Wheel-Running Patterns and PERIOD2 Expression in the Rat Forebrain. *J Mol Neurosci* [Epub ahead of print]

Gray TA, Saitoh S, Nicholls RD (1999a) An imprinted, mammalian bicistronic transcript encodes two independent proteins. *Proc Natl Acad Sci USA* 96(10):5616-5621

Gray TA, Smithwick MJ, Schaldach MA, Martone DL, Graves JA, McCarrey JR, Nicholls RD (1999b) Concerted regulation and molecular evolution of the duplicated SNRPB'/B and SNRPN loci. *Nucleic Acids Res* 27(23):4577-4584

Gray TA, Hernandez L, Carey AH, Schaldach MA, Smithwick MJ, Rus K, Marshall Graves JA, Stewart CL, Nicholls RD (2000) The ancient source of a distinct gene family encoding proteins featuring RING and C(3)H zinc-finger motifs with abundant expression in developing brain and nervous system. *Genomics* 66(1):76-86

Gray TA, Azama K, Whitmore K, Min A, Abe S, Nicholls RD (2001) Phylogenetic conservation of the makorin-2 gene, encoding a multiple zinc-finger protein, antisense to the RAF1 proto-oncogene. *Genomics* 77(3):119-126

Green CB, Douris N, Kojima S, Strayer CA, Fogerty J, Lourim D, Keller SR, Besharse JC (2007) Loss of Nocturnin, a circadian deadenylase, confers resistance to hepatic steatosis and diet-induced obesity. *Proc Natl Acad Sci U S A* 104(23):9888-9893

Griffin EA Jr, Staknis D, Weitz CJ (1999) Light-independent role of CRY1 and CRY2 in the mammalian circadian clock. *Science* 286(5440):768-771

Gugneja S, Virbasius CM, Scarpulla RC (1996) Nuclear respiratory factors 1 and 2 utilize similar glutamine-containing clusters of hydrophobic residues to activate transcription. *Mol Cell Biol* 16(10):5708-5716

Gugneja S, Scarpulla RC (1997) Serine phosphorylation within a concise amino-terminal domain in nuclear respiratory factor 1 enhances DNA binding. *J Biol Chem* 272(30):18732-18739

Ramachandran B, Yu G, Gulick T (2008) Nuclear respiratory factor 1 controls myocyte enhancer factor 2A transcription to provide a mechanism for coordinate expression of respiratory chain subunits. *J Biol Chem* 283(18):11935-46

Gunay-Aygun M, Schwartz S, Heeger S, O'Riordan MA, Cassidy SB (2001) The changing purpose of Prader-Willi syndrome clinical diagnostic criteria and proposed revised criteria. *Pediatrics* 108(5):E92

Hall JG (1990) Genomic imprinting: review and relevance to human disease. *Am. J. Hum. Genet* 46(5):857-873

Hallikas O, Palin K, Sinjushina N, Rautiainen R, Partanen J, Ukkonen E, Taipale J (2006) Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* 124(1):47-59

Hannula K, Lipsanen-Nyman M, Kontiokari T, Kere J (2001) A narrow segment of maternal uniparental disomy of chromosome 7q31-qter in Silver-Russell syndrome delimits a candidate gene region. *Am J Hum Genet* 68(1):247-253

Hara Y, Onishi Y, Oishi K, Miyazaki K, Fukamizu A, Ishida N (2009) Molecular characterization of Mybbp1a as a co-repressor on the Period2 promoter. Nucleic Acids Res 37(4):1115-1126

Hardie DG (2008) AMPK: a key regulator of energy balance in the single cell and the whole organism. *Int J Obes* (Lond) 32 Suppl 4:S7-12

Hastings M, O'Neill JS, Maywood ES (2007) Circadian clocks: regulators of endocrine and metabolic rhythms. *J Endocrinol* 195(2):187-198

Haussmann IU, White K, Soller M (2008) Erect wing regulates synaptic growth in Drosophila by integration of multiple signaling pathways. *Genome Biol* 9(4):R73

He Y, Jones CR, Fujiki N, Xu Y, Guo B, Holder JL Jr, Rossner MJ, Nishino S, Fu YH (2009) The transcriptional repressor DEC2 regulates sleep length in mammals. *Science* 325(5942):866-870

He Z, Hu Y, Feng L, Li Y, Liu G, Xi Y, Wen L, Lucia A (2008) NRF-1 genotypes and endurance exercise capacity in young Chinese men. *Br J Sports Med* 42(5):361-366

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39(3):311-318

Herzig RP, Andersson U, Scarpulla RC (2000) Dynein light chain interacts with NRF-1 and EWG, structurally and functionally related transcription factors from humans and Drosophila. *J Cell Sci* 113 Pt 23:4263-4273

Hintermann E, Grieder NC, Amherd R, Brodbeck D, Meyer UA (1996) Cloning of an arylalkylamine N-acetyltransferase (aaNAT1) from Drosophila melanogaster expressed in the nervous system and the gut. *Proc Natl Acad Sci USA* 93(22):12315-12320

Hirasawa R, Feil R (2010) Genomic imprinting and human disease. *Essays Biochem* 48(1):187-200

Hirayama J, Sahar S, Grimaldi B, Tamaru T, Takamatsu K, Nakahata Y, Sassone-Corsi P. (2007) CLOCK-mediated acetylation of BMAL1 controls circadian function. *Nature* 450(7172):1086-1090

Hirose T, Smith RJ, Jetten AM (1994) ROR gamma: the third member of ROR/RZR orphan receptor subfamily that is highly expressed in skeletal muscle. *Biochem Biophys Res Commun* 205(3):1976-1983

Hirota T, Lee JW, Lewis WG, Zhang EE, Breton G, Liu X, Garcia M, Peters EC, Etchegaray JP, Traver D, Schultz PG, Kay SA (2010) High-throughput chemical screen identifies a novel potent modulator of cellular circadian rhythms and reveals CKIα as a clock regulatory kinase. *PLoS Biol* 8(12):e1000559

Holm VA, Cassidy SB, Butler MG, Hanchett JM, Greenswag LR, Whitman BY, Greenberg F (1993) Prader-Willi syndrome: consensus diagnostic criteria. *Pediatrics* 91(2):398–402

Honma S, Kawamoto T, Takagi Y, Fujimoto K, Sato F, Noshiro M, Kato Y, Honma K (2002) Dec1 and Dec2 are regulators of the mammalian molecular clock. *Nature* 419(6909):841-844

Höög C, Calzone FJ, Cutting AE, Britten RJ, Davidson EH (1991) Gene regulatory factors of the sea urchin embryo. II. Two dissimilar proteins, P3A1 and P3A2, bind to the same target sites that are required for early territorial gene expression. *Development* 112(1):351-364

Horsthemke B, Buiting K (2006) Imprinting defects on human chromosome 15. *Cytogenet Genome Res* 113(1-4):292-299

Huang AM, Wang HL, Tang YP, Lee EH (1998) Expression of integrin-associated protein gene associated with memory formation in rats. *J Neurosci* 18(11):4305-4313

Huang W, Ramsey KM, Marcheva B, Bass J (2011) Circadian rhythms, sleep, and metabolism. *J Clin Invest* 121(6):2133-41. doi: 10.1172/JCI46043

Hughes ME, DiTacchio L, Hayes KR, Vollmers C, Pulivarthy S, Baggs JE, Panda S, Hogenesch JB (2009) Harmonics of circadian gene transcription in mammals. *PLoS Genet* 5(4):e1000442

Huo L, Scarpulla RC (2001) Mitochondrial DNA instability and peri-implantation lethality associated with targeted disruption of nuclear respiratory factor 1 in mice. *Mol Cell Biol* 21(2):644-654

Inoue S, Shimoda M, Nishinokubi I, Siomi MC, Okamura M, Nakamura A, Kobayashi S, Ishida N, Siomi H (2002) A role for the Drosophila fragile X-related gene in circadian output. *Curr Biol* 12(15):1331-1335

Inouye ST, Kawamura H (1979) Persistence of circadian rhythmicity in a mammalian hypothalamic "island" containing the suprachiasmatic nucleus. *Proc Natl Acad Sci USA* 76(11):5962-5966

Izumi H, Ohta R, Nagatani G, Ise T, Nakayama Y, Nomoto M, Kohno K. (2003) p300/CBP-associated factor (P/CAF) interacts with nuclear respiratory factor-1 to regulate the UDP-N-acetyl-alpha-dgalactosamine: polypeptide N-acetylgalactosaminyltransferase-3 gene. *Biochem J* 373(Pt 3):713-722

Jay P, Rougeulle C, Massacrier A, Moncla A, Mattei MG, Malzac P, Roeckel N, Taviaux S, Lefranc JL, Cau P, Berta P, Lalande M, Muscatelli F (1997) The human necdin gene, NDN, is maternally imprinted and located in the Prader-Willi syndrome chromosomal region. *Nature Genet* 17(3):357-361

Jones AR, Overly CC, Sunkin SM (2009) The Allen Brain Atlas: 5 years and beyond. *Nat Rev Neurosci* 10(11):821-828

Jong MT, Carey AH, Caldwell KA, Lau MH, Handel MA, Driscoll DJ, Stewart CL, Rinchik EM, Nicholls RD (1999a) Imprinting of a RING zinc-finger encoding gene in the mouse chromosome region homologous to the Prader-Willi syndrome genetic region. *Hum Mol Genet* 8(5):795-803

Jong MT, Gray TA, Ji Y, Glenn CC, Saitoh S, Driscoll DJ, Nicholls RD (1999b) A novel imprinted gene, encoding a RING zinc-finger protein, and overlapping antisense transcript in the Prader-Willi syndrome critical region. *Hum Mol Genet* 8(5):783-793

Kaasik K, Lee CC (2004) Reciprocal regulation of haem biosynthesis and the circadian clock in mammals. *Nature* 430(6998):467-471

Kapfhamer D, Valladares O, Sun Y, Nolan PM, Rux JJ, Arnold SE, Veasey SC, Bućan M (2002) Mutations in Rab3a alter circadian period and homeostatic response to sleep loss in the mouse. *Nat Genet* 32(2):290-295

Katada S, Sassone-Corsi P (2010) The histone methyltransferase MLL1 permits the oscillation of circadian gene expression. *Nat Struct Mol Biol* 17(12):1414-1421

Kayashima T, Katahira M, Harada N, Miwa N, Ohta T, Yoshiura K, Matsumoto N, Nakane Y, Nakamura Y, Kajii T, Niikawa N, Kishino T (2002) Maternal isodisomy for 14q21-q24 in a man with diabetes mellitus. *Am J Med Genet* 111(1):38-42

Keller U (2006) From obesity to diabetes. *Int J Vitam Nutr Res* 76(4):172-177

Kelly DP, Scarpulla RC (2004) Transcriptional regulatory circuits controlling mitochondrial biogenesis and function. *Genes Dev* 18(4):357-368

Khan MS, Maden BE (1976) Nucleotide sequences within the ribosomal ribonucleic acids of HeLa cells, Xenopus laevis and chick embryo fibroblasts. *J Mol Biol* 101(2):235-254

Killian JK, Nolan CM, Wylie AA, Li T, Vu TH, Hoffman AR, Jirtle RL (2001) Divergent evolution in M6P/IGF2R imprinting from the Jurassic to the Quaternary. *Hum Mol Genet* 10(17):1721-1728

Kim SA, Yoon JH, Lee SH, Ahn SG (2005) Polo-like kinase 1 phosphorylates heat shock transcription factor 1 and mediates its nuclear translocation during heat stress. *J Biol Chem* 280(13):12653-12657

Kim JD, Yu S, Kim J (2009) YY1 is autoregulated through its own DNA-binding sites. *BMC Mol Biol* 10:85

Kishore S, Stamm S (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* 311(5758):230-232

Kiss T (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* 109(2):145-148

Kiss-László Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* 85(7): 1077-1088

Kiss-László Z, Henry Y, Kiss T (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J* 17(3):797-807

Kiyohara YB, Nishii K, Ukai-Tadenuma M, Ueda HR, Uchiyama Y, Yagita K (2008) Detection of a circadian enhancer in the mDbp promoter using prokaryotic transposon vector-based strategy. *Nucleic Acids Res* 36(4):e23

Koh K, Zheng X, Sehgal A (2006) JETLAG resets the *Drosophila* circadian clock by promoting light-induced degradation of TIMELESS. *Science* 312(5781):1809-1812

Kondratov RV, Chernov MV, Kondratova AA, Gorbacheva VY, Gudkov AV, Antoch MP (2003) BMAL1-dependent circadian oscillation of nuclear CLOCK: posttranslational events induced by dimerization of transcriptional activators of the mammalian clock system. *Genes Dev* 17(15):1921-1932

Kondratov RV, Kondratova AA, Lee C, Gorbacheva VY, Chernov MV, Antoch MP (2006) Post-translational regulation of circadian transcriptional CLOCK(NPAS2)/BMAL1 complex by CRYPTOCHROMES. *Cell Cycle* 5(8):890-895

Kondratov RV, Antoch MP (2007) The clock proteins, aging, and tumorigenesis. *Cold Spring Harb Symp Quant Biol* 72:477-482

Kota SK, Feil R (2010) Epigenetic transitions in germ cell development and meiosis. Dev *Cell* 19(5):675-686

Kouris-Blazos A, Wahlqvist ML (2007) Health economics of weight management: evidence and cost. *Asia Pac J Clin Nutr* 16 Suppl 1:329-338

Kozlov SV, Bogenpohl JW, Howell MP, Wevrick R, Panda S, Hogenesch JB, Muglia LJ, Van Gelder RN, Herzog ED, Stewart CL (2007) The imprinted gene Magel2 regulates normal circadian output. *Nat Genet* 39(10):1266-1272

Krol J, Busskamp V, Markiewicz I, Stadler MB, Ribi S, Richter J, Duebel J, Bicker S, Fehling HJ, Schübeler D, Oertner TG, Schratt G, Bibel M, Roska B, Filipowicz W (2010) Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. *Cell* 141(4):618-631

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. *Nature* 392(6679): 917-920

Kuo MH, Allis CD (1999) In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* 19(3):425-433

Lamia KA, Sachdeva UM, DiTacchio L, Williams EC, Alvarez JG, Egan DF, Vasquez DS, Juguilon H, Panda S, Shaw RJ and others (2009) AMPK regulates the circadian clock by cryptochrome phosphorylation and degradation. *Science* 326(5951):437-440

Landschulz WH, Johnson PF, Adashi EY, Graves BJ, McKnight SL (1988) Isolation of a recombinant copy of the gene encoding C/EBP. *Genes Dev* 2(7):786-800

Langmesser S, Tallone T, Bordon A, Rusconi S, Albrecht U (2008) Interaction of circadian clock proteins PER2 and CRY with BMAL1 and CLOCK. *BMC Mol Biol* 9:41

Larsson E, Lindahl P, Mostad P (2007) HeliCis: a DNA motif discovery tool for colocalized motif pairs with periodic spacing. *BMC Bioinformatics* 8:418

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11(3):204-220

Lawton BR, Carone BR, Obergfell CJ, Ferreri GC, Gondolphi CM, Vandeberg JL, Imumorin I, O'Neill RJ, O'Neill MJ (2008) Genomic imprinting of IGF2 in marsupials is methylation dependent. *BMC Genomics* 9:205

Lee C, Etchegaray JP, Cagampang FR, Loudon AS, Reppert SM (2001) Posttranslational mechanisms regulate the mammalian circadian clock. *Cell* 107(7):855-867

Lee H, Chen R, Lee Y, Yoo S, Lee C (2009) Essential roles of CKIdelta and CKIepsilon in the mammalian circadian clock. *Proc Natl Acad Sci U S A* 106(50):21359-21364

Lee S, Kozlov S, Hernandez L, Chamberlain SJ, Brannan CI, Stewart CL, Wevrick R (2000) Expression and imprinting of MAGEL2 suggest a role in Prader-Willi syndrome and the homologous murine imprinting phenotype. *Hum Mol Genet* 9(12):1813-1819

Lee S, Walker CL, Wevrick R (2003) Prader-Willi syndrome transcripts are expressed in phenotypically significant regions of the developing mouse brain. *Gene Expr Patterns* 3(5):599-609

Lee S, Walker CL, Karten B, Kuny SL, Tennese AA, O'Neill MA, Wevrick R (2005) Essential role for the Prader-Willi syndrome protein necdin in axonal outgrowth. *Hum Mol Genet* 14(5):627-637

Lemos DR, Goodspeed L, Tonelli L, Antoch MP, Ojeda SR, Urbanski HF (2007) Evidence for circadian regulation of activating transcription factor 5 but not tyrosine hydroxylase by the chromaffin cell clock. *Endocrinology* 148(12):5811-5821

Levi F, Schibler U (2007) Circadian rhythms: mechanisms and therapeutic implications. *Annu Rev Pharmacol Toxicol* 47:593-628

Lezza AM, Pesce V, Cormio A, Fracasso F, Vecchiet J, Felzani G, Cantatore P, Gadaleta MN (2001) Increased expression of mitochondrial transcription factor A and nuclear respiratory factor-1 in skeletal muscle from aged human subjects. *FEBS Lett* 501(1):74-78

Li Y, Xie M, Song X, Gragen S, Sachdeva K, Wan Y, Yan B (2003) DEC1 negatively regulates the expression of DEC2 through binding to the E-box in the proximal promoter. *J Biol Chem* 278(19):16899-16907

Li Y, Sasaki H (2011) Genomic imprinting in mammals: its life cycle, molecular mechanisms and reprogramming. *Cell Res* 21(3):466-73

Liu C, Li S, Liu T, Borjigin J, Lin JD (2007) Transcriptional coactivator PGC-1alpha integrates the mammalian clock and energy metabolism. *Nature* 447(7143):477-481

Liang HL, Wong-Riley MTT (2006) Activity-dependent regulation of nuclear respiratory factor-I, nuclear respiratory factor-2, and peroxisome proliferator-activated receptor gamma coactivator-I in neurons. *Neuroreport* 17(4):401-405

Lim CT, Kola B, Korbonits M (2010) AMPK as a mediator of hormonal signaling *J Mol Endocrinol* 44(2):87-97

Liu AC, Tran HG, Zhang EE, Priest AA, Welsh DK, Kay SA (2008) Redundant function of REV-ERBalpha and beta and non-essential role for Bmal1 cycling in transcriptional regulation of intracellular circadian rhythms. *PLoS Genet* 4(2):e1000023

Liu C, Li S, Liu T, Borjigin J, Lin JD (2007) Transcriptional coactivator PGC-1alpha integrates the mammalian clock and energy metabolism. *Nature* 447(7143):477-481

Liu Y, Niu N, Zhu X, Du T, Wang X, Chen D, Wu X, Gu HF, Liu Y (2008) Genetic variation and association analyses of the nuclear respiratory factor 1 (nRF1) gene in Chinese patients with type 2 diabetes. *Diabetes* 57(3):777-782

Lopez-Molina L, Conquet F, Dubois-Dauphin M, Schibler U (1997) The DBP gene is expressed according to a circadian rhythm in the suprachiasmatic nucleus and influences circadian behavior. *EMBO J* 16(22):6762-6771

Lowell BB, Shulman GI (2005) Mitochondrial dysfunction and type 2 diabetes. *Science* 307(5708):384-387

MacDonald HR, Wevrick R (1997) The necdin gene is deleted in Prader-Willi syndrome and is imprinted in human and mouse. *Hum Mol Genet* 6(11):1873-1878

MacQuarrie KL, Fong AP, Morse RH, Tapscott SJ (2011) Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet* 27(4):141-148

Maden BE. (1990) The numerous modified nucleotides in eukaryotic ribosomal RNA. Prog *Nucleic Acid Res Mol Biol* 39:241-303

Maden BE, Hughes JM (1997) Eukaryotic ribosomal RNA: the recent excitement in the nucleotide modification problem. *Chromosoma* 105(7-8):391-400

Maden BE (2001) Mapping 2'-O-methyl groups in ribosomal RNA. *Methods* 25(3):374-382

Mahony S, Benos PV (2007) STAMP: a web tool for exploring DNA-binding motif similarities. *Nucl Acids Res* 35 (Web Server issue): W253-258

Marcheva B, Ramsey KM, Buhr ED, Kobayashi Y, Su H, Ko CH, Ivanova G, Omura C, Mo S, Vitaterna MH, Lopez JP, Philipson LH, Bradfield CA, Crosby SD, JeBailey L, Wang X, Takahashi JS, Bass J (2010) Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes *Nature* 466(7306):627-631

Masri S, Sassone-Corsi P (2010) Plasticity and specificity of the circadian epigenome. *Nat Neurosci* 13(11):1324-1329

Matera AG, Terns RM, Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* 8(3):209-220

Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288(5):911-940

Macneil LT, Walhout AJ (2011) Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res* 21(5):645-657

Maywood ES, O'Neill JS, Chesham JE, Hastings MH (2007) The circadian clockwork of the suprachiasmatic nuclei--analysis of a cellular oscillator that drives endocrine rhythms. *Endocrinology* 148(12):5624-5634

McAllister G, Roby-Shemkovitz A, Amara SG, Lerner MR (1989) cDNA sequence of the rat U snRNP-associated protein N: description of a potential Sm epitope. EMBO J 8(4):1177-81

McDaniel LD, Schultz RA (1992) Elevated sister chromatid exchange phenotype of Bloom syndrome cells is complemented by human chromosome 15. *Proc Natl Acad Sci USA* 89(17):7968-7972

Mendoza J (2007) Circadian clocks: setting time by food. *J Neuroendocrinol* 19(2):127-137

Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ,Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SM, Sharpe T, Baker ML, Batzer MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidow L, Deakin JE, Fazzari MJ,Glass JL, Grabherr M, Greally JM, Gu W, Hore TA, Huttley GA, Kleber M, Jirtle RL, Koina E,Lee JT, Mahony S, Marra MA, Miller RD, Nicholls RD, Oda M, Papenfuss AT, Parra ZE, Pollock DD, Ray DA, Schein JE, Speed TP, Thompson K, VandeBerg JL, Wade CM, Walker JA, Waters PD, Webber C, Weidman JR, Xie X, Zody MC; Broad Institute Genome Sequencing Platform;Broad Institute Whole Genome Assembly Team, Graves JA, Ponting CP, Breen M, Samollow PB, Lander ES, Lindblad-Toh K (2007) Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. Nature 447(7141):167-177

Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet* 34(3):267-273

Morison IM, Paton CJ, Cleverley SD (2001) The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res* 29(1):275-276

Morison IM, Ramsay JP, Spencer HG (2005) A census of mammalian imprinting. *Trends Genet* 21(8):457-465

Morrish F, Giedt C, Hockenbery D (2003) c-MYC apoptotic function is mediated by NRF-1 target genes. *Genes Dev* 17(2):240-255

Motzkus D, Loumi S, Cadenas C, Vinson C, Forssmann WG, Maronde E (2007) Activation of human period-1 by PKA or CLOCK/BMAL1 is conferred by separate signal transduction pathways. *Chronobiol Int* 24(5):783-792

Murakami T, Shimomura Y, Yosimura A, Sokabe M, Fujitsuka N (1998) Induction of nuclear respiratory factor-1 expression by an acute bout of exercise in rat muscle. *Biochim Biophys Acta* 1381(1):113–122

Nakahata Y, Kaluzova M, Grimaldi B, Sahar S, Hirayama J, Chen D, Guarente LP, Sassone-Corsi P (2008) The NAD(+)-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control. *Cell* 134(2):329-340

Naruse Y, Oh-hashi K, Iijima N, Naruse M, Yoshioka H, Tanaka M (2004) Circadian and light-induced transcription of clock gene Per1 depends on histone acetylation and deacetylation. *Mol Cell Biol* 24(14):6278-6287

Nicholls RD, Knoll JH, Butler MG, Karam S, Lalande M (1989) Genetic imprinting suggested by maternal heterodisomy in nondeletion Prader-Willi syndrome. *Nature* 342(6247):281-285

Nicholls RD, Knepper JL (2001) Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annu Rev Genomics Hum Genet* 2:153-175

Ning Y, Lovell M, Taylor L, Pereira-Smith OM (1992) Isolation of monochromosomal hybrids following fusion of human diploid fibroblast-derived microcells with mouse A9 cells. *Cytogenet Cell Genet* 60(1):79-80

Nolte C, Rastegar M, Amores A, Bouchard M, Grote D, Maas R, Kovacs EN, Postlethwait J, Rambaldi I, Rowan S, Yan YL, Zhang F, Featherstone M (2006) Stereospecificity and PAX6 function direct Hoxd4 neural enhancer activity along the antero-posterior axis. *Dev Biol* 299(2):582-593

Noshiro M, Kawamoto T, Furukawa M, Fujimoto K, Yoshida Y, Sasabe E, Tsutsumi S, Hamada T, Honma S, Honma K, Kato Y (2004) Rhythmic expression of DEC1 and DEC2 in peripheral tissues: DEC2 is a potent suppressor for hepatic cytochrome P450s opposing DBP. *Genes Cells* 9(4):317-329

Numakawa T, Ishimoto T, Suzuki S, Numakawa Y, Adachi N, Matsumoto T, Yokomaku D, Koshimizu H, Fujimori KE, Hashimoto R, Taguchi T, Kunugi H (2004) Neuronal roles of the integrin-associated protein (IAP/CD47) in developing cortical neurons. *J Biol Chem* 279(41):43245-43253

Ogata T, Kagami M, Ferguson-Smith AC (2008) Molecular mechanisms regulating phenotypic outcome in paternal and maternal uniparental disomy for chromosome 14. *Epigenetics* 3(4):181-187

Ohdo S (2010) Chronotherapeutic strategy: Rhythm monitoring, manipulation and disruption. *Adv Drug Deliv Rev* 62(9-10):859-875

Ohno T, Onishi Y, Ishida N (2007) A novel E4BP4 element drives circadian expression of mPeriod2. *Nucleic Acids Res* 35(2):648-655

Ojuka EO (2004) Role of calcium and AMP kinase in the regulation of mitochondrial biogenesis and GLUT4 levels in muscle. *Proc Nutr Soc* 63(2):275-278

O'Neill JS (2009) Circadian clocks can take a few transcriptional knocks. *EMBO J* 28(2):84-85

O'Neill MJ, Ingram RS, Vrana PB, Tilghman SM (2000) Allelic expression of IGF2 in marsupials and birds. *Dev Genes Evol* 210(1):18-20

Osland TM, Fernø J, Håvik B, Heuch I, Ruoff P, Lærum OD, Steen VM (2010) Lithium Differentially Affects Clock Gene Expression in Serum-Shocked nih-3t3 Cells. *J Psychopharmacol* [Epub ahead of print]

Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M, Schultz PG, Kay SA, Takahashi JS, Hogenesch JB (2002) Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* 109(3):307-320

Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. *J Appl Genet* [Epub ahead of print]

Park SH, Zhu PP, Parker RL, Blackstone C (2010) Hereditary spastic paraplegia proteins REEP1, spastin, and atlastin-1 coordinate microtubule interactions with the tubular ER network. *J Clin Invest* 120(4):1097-1110

Partch CL, Shields KF, Thompson CL, Selby CP, Sancar A (2006) Posttranslational regulation of the mammalian circadian clock by cryptochrome and protein phosphatase 5. *Proc Natl Acad Sci U S A* 103(27):10467-10472

Patti ME, Butte AJ, Crunkhorn S, Cusi K, Berria R, Kashyap S, Miyazaki Y, Kohane I, Costello M, Saccone R, Landaker EJ, Goldfine AB, Mun E, DeFronzo R, Finlayson J, Kahn CR, Mandarino LJ (2003) Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proc Natl Acad Sci U S A* 100(14):8466-8471

Pei DQ, Shih CH (1990) Transcriptional activation and repression by cellular DNA-binding protein C/EBP. *J Virol* 64(4):1517-1522

Porter AC, Itzhaki JE (1993) Gene targeting in human somatic cells. Complete inactivation of an interferon-inducible gene. *Eur J Biochem* 218(2):273-281

Preitner N, Damiola F, Lopez-Molina L, Zakany J, Duboule D, Albrecht U, Schibler U (2002) The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* 110(2):251-260

Pugh BF, Tjian R (1990) Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* 61(7):1187-1197

Raghuram S, Stayrook KR, Huang P, Rogers PM, Nosie AK, McClure DB, Burris LL, Khorasanizadeh S, Burris TP, Rastinejad F (2007) Identification of heme as the ligand for the orphan nuclear receptors REV-ERBalpha and REV-ERBbeta. *Nat Struct Mol Biol* 14(12):1207-1213

Rapkins RW, Hore T, Smithwick M, Ager E, Pask AJ, Renfree MB, Kohn M, Hameister H, Nicholls RD, Deakin JE, Graves JA (2006) Recent assembly of an imprinted domain from non-imprinted components. *PLoS Genet* 2(10):e182

Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes. *RNA* 10(10):1507-1517

Reinke H, Saini C, Fleury-Olela F, Dibner C, Benjamin IJ, Schibler U (2008) Differential display of DNA-binding proteins reveals heat-shock factor 1 as a circadian transcription factor. *Genes Dev* 22(3):331-345

Reischl S, Vanselow K, Westermark PO, Thierfelder N, Maier B, Herzel H, Kramer A (2007) Beta-TrCP1-mediated degradation of PERIOD2 is essential for circadian dynamics. *J Biol Rhythms* 22(5):375-386

Ren J, Lee S, Pagliardini S, Gérard M, Stewart CL, Greer JJ, Wevrick R (2003) Absence of Ndn, encoding the Prader-Willi syndrome-deleted gene necdin, results in congenital deficiency of central respiratory drive in neonatal mice. *J Neurosci* 23(5):1569-1573

Renfree MB (2007) The genome of a Gondwanan mammal. *Bioessays* 29(11):1073-1076

Renfree MB, Hore TA, Shaw G, Graves JA, Pask AJ (2009) Evolution of genomic imprinting: insights from marsupials and monotremes. *Annu Rev Genomics Hum Genet* 10:241-262

Reppert SM, Weaver DR (2002) Coordination of circadian timing in mammals. *Nature* 418(6901):935-941

Richter EA, Ruderman NB (2009) AMPK and the biochemistry of exercise: implications for human health and disease. *Biochem J* 418(2):261-275

Riethoven JJ (2010) Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. *Methods Mol Biol* 674:33-42

Ripperger JA, Shearman LP, Reppert SM, Schibler U (2000) CLOCK, an essential pacemaker component, controls expression of the circadian transcription factor DBP. *Genes Dev* 14(6):679-689

Ripperger JA, Schibler U (2006) Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian Dbp transcription and chromatin transitions. *Nat Genet* 38(3):369-374

Rivera RM, Bennett LB (2010) Epigenetics in humans: an overview. *Curr Opin Endocrinol Diabetes Obes* 17(6):493-499

Robinson WP, Christian SL, Kuchinka BD, Peñaherrera MS, Das S, Schuffenhauer S, Malcolm S, Schinzel AA, Hassold TJ, Ledbetter DH (2000) Somatic segregation errors predominantly contribute to the gain or loss of a paternal chromosome leading to uniparental disomy for chromosome 15. *Clin Genet* 57(5):349-358

Rodriguez-Jato S, Nicholls RD, Driscoll DJ, Yang TP (2005) Characterization of cis-and trans-acting elements in the imprinted human SNURF-SNRPN locus. *Nucl Acids Res* 33(15):4740-4753

Roenneberg T, Merrow M (2005) Circadian clocks - the fall and rise of physiology. *Nat Rev Mol Cell Biol* 6(12):965-971

Runte M, Hüttenhofer A, Gross S, Kiefmann M, Horsthemke B, Buiting K (2001) The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum Mol Genet* 10(23):2687-2700

Rutter J, Reick M, McKnight SL (2002) Metabolism and the control of circadian rhythms. *Annu Rev Biochem* 71:307-331

Salafsky IS, MacGregor SN, Claussen U, von Eggeling F (2001) Maternal UPD 20 in an infant from a pregnancy with mosaic trisomy 20. *Prenat Diagn* 21(10):860-863

Salinas S, Proukakis C, Crosby A, Warner TT (2008) Hereditary spastic paraplegia: clinical features and pathogenetic mechanisms. *Lancet Neurol* 7(12):1127-1138

Sasaki M, Yoshitane H, Du NH, Okano T, Fukada Y (2009) Preferential inhibition of BMAL2-CLOCK activity by PER2 reemphasizes its negative role and a positive role of BMAL2 in the circadian transcription. *J Biol Chem* 284(37):25149-25159

Sato F, Kawamoto T, Fujimoto K, Noshiro M, Honda KK, Honma S, Honma K, Kato Y (2004) Functional analysis of the basic helix-loop-helix transcription factor DEC1 in circadian regulation. Interaction with BMAL1. *Eur J Biochem* 271(22):4409-4419

Scarpulla RC (2006) Nuclear control of respiratory gene expression in mammalian cells. *J Cell Biochem* 97(4):673-683

Scarpulla RC (2008) Transcriptional paradigms in mammalian mitochondrial biogenesis and function. *Physiol Rev* 88(2):611-638

Schübeler D (2007) Enhancing genome annotation with chromatin. *Nat Genet* 39(3):284-285

Schuetz A, Allali-Hassani A, Martín F, Loppnau P, Vedadi M, Bochkarev A, Plotnikov AN, Arrowsmith CH, Min J (2006) Structural basis for molecular recognition and presentation of histone H3 by WDR5. *EMBO J* 25(18):4245-4252

Sedivy JM, Dutriaux A (1999) Gene targeting and somatic cell genetics--a rebirth or a coming of age? *Trends Genet* 15(3):88-90

Sharpe NG, Williams DG, Latchman DS (1990) Regulated expression of the small nuclear ribonucleoprotein particle protein SmN in embryonic stem cell differentiation. *Mol Cell Biol* 10(12):6817-6820

Shi S, Hida A, McGuinness OP, Wasserman DH, Yamazaki S, Johnson CH (2010) Circadian clock gene Bmal1 is not essential; functional replacement with its paralog, Bmal2. *Curr Biol* 20(4):316-321

Shield JP (2000) Neonatal diabetes: new insights into aetiology and implications. *Horm Res* 53 Suppl 1:7-11

Shim HS, Kim H, Lee J, Son GH, Cho S, Oh TH, Kang SH, Seen DS, Lee KH, Kim K (2007) Rapid activation of CLOCK by Ca2+-dependent protein kinase C mediates resetting of the mammalian circadian clock. *EMBO Rep* 8(4):366-371

Shirogane T, Jin J, Ang XL, Harper JW (2005) SCFbeta-TRCP controls clock-dependent transcription via casein kinase 1-dependent degradation of the mammalian period-1 (Per1) protein. *J Biol Chem* 280(29):26863-26872

Smith AC, Shuman C, Chitayat D, Steele L, Ray PN, Bourgeois J, Weksberg R (2007) Severe presentation of Beckwith-Wiedemann syndrome associated with high levels of

constitutional paternal uniparental disomy for chromosome 11p15. *Am J Med Genet A* 143A(24):3010-3015

Smith KT, Coffee B, Reines D (2004) Occupancy and synergistic activation of the FMR1 promoter by Nrf-1 and Sp1 in vivo. *Hum Mol Genet* 13(15):1611-1621

Solecki D, Bernhardt G, Lipp M, Wimmer E (2000) Identification of a nuclear respiratory factor-1 binding site within the core promoter of the human polio virus receptor/CD155 gene. *J Biol Chem* 275(17):12453-12462

Spasic MR, Callaerts P, Norga KK (2009) AMP-activated protein kinase (AMPK) molecular crossroad for metabolic control and survival of neurons. *Neuroscientist* 15(4):309-316

Spitz F, Duboule D (2008) Chapter 6 global control regions and regulatory landscapes in vertebrate development and evolution. *Adv Genet* 61:175-205

Spengler ML, Kuropatwinski KK, Schumer M, Antoch MP. (2009) A serine cluster mediates BMAL1-dependent CLOCK phosphorylation and degradation. *Cell Cycle* 8(24):4138-4146

Steeves TD, King DP, Zhao Y, Sangoram AM, Du F, Bowcock AM, Moore RY, Takahashi JS (1999) Molecular cloning and characterization of the human CLOCK gene: expression in the suprachiasmatic nuclei. *Genomics* 57(2):189-200

Stefan M, Ji H, Simmons RA, Cummings DE, Ahima RS, Friedman MI, Nicholls RD (2005) Hormonal and metabolic defects in a prader-willi syndrome mouse model with neonatal failure to thrive. *Endocrinology* 146(10): 4377-4385

Stefan M, Simmons RA, Bertera S, Trucco M, Esni F, Drain P, Nicholls RD (2011) Global deficits in development, function, and gene expression in the endocrine pancreas in a deletion mouse model of Prader-Willi syndrome. *Am J Physiol Endocrinol Metab* 300(5):E909-922

Stefan M, Ohta T, **Zhu W**, Yamasaki K, Benos PV, Nicholls RD. Nuclear respiratory factor-1 is a master transcriptional regulator over a 2 Mb mammalian chromosomal domain. *Manuscript in preparation*

Steeves TD, King DP, Zhao Y, Sangoram AM, Du F, Bowcock AM, Moore RY, Takahashi JS (1999) Molecular cloning and characterization of the human CLOCK gene: expression in the suprachiasmatic nuclei. *Genomics* 57(2):189-200

Stehle JH, Foulkes NS, Molina CA, Simonneaux V, Pévet P, Sassone-Corsi P (1993) Adrenergic signals direct rhythmic expression of transcriptional repressor CREM in the pineal gland. *Nature* 365(6444):3143-20

Stratmann M, Stadler F, Tamanini F, van der Horst GT, Ripperger JA (2010) Flexible phase adjustment of circadian albumin D site-binding protein (DBP) gene expression by CRYPTOCHROME1. *Genes Dev* 24(12):1317-1328

Sun ZS, Albrecht U, Zhuchenko O, Bailey J, Eichele G, Lee CC (1997) RIGUI, a putative mammalian ortholog of the *Drosophila* period gene. *Cell* 90(6):1003-1011

Surani MA, Barton SC, Norris ML (1986) Nuclear transplantation in the mouse: heritable differences between parental genomes after activation of the embryonic genome. *Cell* 45(1):127-136

Suter DM, Schibler U (2009) Feeding the clock. *Science* 326(5951):378-379

Suzuki S, Renfree MB, Pask AJ, Shaw G, Kobayashi S, Kohda T, Kaneko-Ishino T, Ishino F (2005) Genomic imprinting of IGF2, p57(KIP2) and PEG1/MEST in a marsupial, the tammar wallaby. *Mech Dev* 122(2):213-222

Suzuki S, Ono R, Narita T, Pask AJ, Shaw G, Wang C, Kohda T, Alsop AE, Marshall Graves JA, Kohara Y, Ishino F, Renfree MB, Kaneko-Ishino T (2007) Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. *PLoS Genet* 3(4):e55

Takahashi JS, Hong HK, Ko CH, McDearmon EL (2008) The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nature Rev Genet* 9(10):764-775

Tamaru T, Isojima Y, van der Horst GT, Takei K, Nagai K, Takamatsu K (2003) Nucleocytoplasmic shuttling and phosphorylation of BMAL1 are regulated by circadian clock in cultured fibroblasts. *Genes Cells* 8(12):973-983

Tanaka-Fujita R, Soeno Y, Satoh H, Nakamura Y, Mori S (2007) Human and mouse protein-noncoding snoRNA host genes with dissimilar nucleotide sequences show chromosomal synteny. *RNA* 13(6):811-816

Timmons JA, Norrbom J, Schéele C, Thonberg H, Wahlestedt C, Tesch P (2006) Expression profiling following local muscle inactivity in humans provides new perspective on diabetes-related genes. *Genomics* 87:165-172

Toh KL (2008) Basic science review on circadian rhythm biology and circadian sleep disorders. *Ann Acad Med Singapore* 37(8):662-668

Triqueneaux G, Thenot S, Kakizawa T, Antoch MP, Safi R, Takahashi JS, Delaunay F, Laudet V (2004) The orphan receptor Rev-erbalpha gene is a target of the circadian clock pacemaker. *J Mol Endocrinol* 33(3):585-608

Tsai TF, Jiang YH, Bressler J, Armstrong D, Beaudet AL (1999) Paternal deletion from Snrpn to Ube3a in the mouse causes hypotonia, growth retardation and partial lethality and

provides evidence for a gene contributing to Prader-Willi syndrome. *Hum Mol Genet* 8(8):1357-1364

Tseng YH, Butte AJ, Kokkotou E, Yechoor VK, Taniguchi CM, Kriauciunas KM, Cypess AM, Niinobe M, Yoshikawa K, Patti ME, Kahn CR (2005) Prediction of preadipocyte differentiation by gene expression reveals role if insulin receptor substrates and necdin. *Nature Cell Biol* 7(6):601-611

Tsuchiya Y, Akashi M, Matsuda M, Goto K, Miyata Y, Node K, Nishida E (2009) Involvement of the protein kinase CK2 in the regulation of mammalian circadian rhythms. *Sci Signal* 2(73):ra26

Turek FW, Joshu C, Kohsaka A, Lin E, Ivanova G, McDearmon E, Laposky A, Losee-Olson S, Easton A, Jensen DR, Eckel RH, Takahashi JS, Bass J (2005) Obesity and metabolic syndrome in circadian Clock mutant mice. *Science* 308(5724):1043-1045

Tyc K, Steitz JA (1989) U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J* 8(10):3113-3119

Tycowski KT, Shu MD, Steitz JA (1996) A mammalian gene with introns instead of exons generating stable RNA products. *Nature* 379(6564):464-466

Ueda HR, Chen W, Adachi A, Wakamatsu H, Hayashi S, Takasugi T, Nagano M, Nakahama K,Suzuki Y, Sugano S, Iino M, Shigeyoshi Y, Hashimoto S (2002) A transcription factor response element for gene expression during circadian night. *Nature* 418(6897):534-539

Ueda HR, Hayashi S, Chen W, Sano M, Machida M, Shigeyoshi Y, Iino M, Hashimoto S (2005) System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nat Genet* 37(2):187-192

Unsal-Kaçmaz K, Mullen TE, Kaufmann WK, Sancar A (2005) Coupling of human circadian and cell cycles by the timeless protein. *Mol Cell Biol* 25(8):3109-3116

Uribe-Lewis S, Woodfine K, Stojic L, Murrell A (2011) Molecular mechanisms of genomic imprinting and clinical implications for cancer. *Expert Rev Mol Med* 13:e2

Vallone D, Lahiri K, Dickmeis T, Foulkes NS (2007) Start the clock! Circadian rhythms and development. *Dev Dyn* 236(1):142-155

van Rheede T, Bastiaans T, Boone DN, Hedges SB, de Jong WW, Madsen O (2006) The platypus is in its place: nuclear genes and indels confirm the sister group relation of monotremes and Therians. *Mol Biol Evol* 23(3):587-597

Vilen BJ, Cogswell JP, Ting JP (1991) Stereospecific alignment of the X and Y elements is required for major histocompatibility complex class II DRA promoter function. *Mol Cell Biol* 11(5):2406-2415

Villard J (2004) Transcription regulation and human diseases. *Swiss Med Wkly* 134(39-40):571-579

Viña J, Gomez-Cabrera MC, Borras C, Froio T, Sanchis-Gomar F, Martinez-Bello VE, Pallardo FV (2009) Mitochondrial biogenesis in exercise and in ageing. *Adv Drug Deliv Rev* 61(14):1369-1374

Virbasius CA, Virbasius JV, Scarpulla RC (1993) NRF-1, an activator involved in nuclear-mitochondrial interactions, utilizes a new DNA-binding domain conserved in a family of developmental regulators. *Genes Dev* 7(12A):2431-2445

Virbasius JV, Scarpulla RC (1994) Activation of the human mitochondrial transcription factor A gene by nuclear respiratory factors: a potential regulatory link between nuclear and mitochondrial gene expression in organelle biogenesis. *Proc Natl Acad Sci USA* 91(4):1309-1313

von Schantz M, Jenkins A, Archer SN (2006) Evolutionary history of the vertebrate period genes. *J Mol Evol* 62(6):701-707

Wang JL, Chang WT, Tong CW, Kohno K, Huang AM (2009) Human synapsin I mediates the function of nuclear respiratory factor 1 in neurite outgrowth in neuroblastoma IMR-32 cells. *J Neurosci Res* 87(10):2255-2263

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, Chang HY (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472(7341):120-124

Wasserman WW, Palumbo M, Thompson W, Fickett JW, Lawrence CE (2000) Human-mouse genome comparisons to locate regulatory sites. *Nat Genet* 26(2):225-228

Wasserman WW, Sandelin A (2004) Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet* 5(4):276-287

Watrin F, Le Meur E, Roeckel N, Ripoche MA, Dandolo L, Muscatelli F (2005) The Prader-Willi syndrome murine imprinting center is not involved in the spatio-temporal transcriptional regulation of the Necdin gene. *BMC Genet* 6:1

Wattendorf DJ, Muenke M (2005) Prader-Willi syndrome. *Am Fam Physician* 72(5):827-830

Weber MJ (2006) Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet* 2(12):e205

Weidman JR, Dolinoy DC, Maloney KA, Cheng JF, Jirtle RL (2006) Imprinting of opossum Igf2r in the absence of differential methylation and air. *Epigenetics* 1(1):49-54

Weldemichael DA, Grossberg GT (2010) Circadian rhythm disturbances in patients with Alzheimer's disease: a review. *Int J Alzheimers Dis* pii: 716453

Wevrick R, Kerns JA, Francke U (1994) Identification of a novel paternally expressed gene in the Prader-Willi syndrome region. *Hum Mol Genet* 3(10):1877-1882

Wirth J, Back E, Hüttenhofer A, Nothwang HG, Lich C, Gross S, Menzel C, Schinzel A, Kioschis P, Tommerup N, Ropers HH, Horsthemke B, Buiting K (2001) A translocation breakpoint cluster disrupts the newly defined 3' end of the SNURF-SNRPN transcription unit on chromosome 15. *Hum Mol Genet* 10(3):201-210

Wood AJ, Oakey RJ (2006) Genomic imprinting in mammals: emerging themes and established theories. *PLoS Genet* 2(11):e147

Wright DC, Han D-H, Garcia-Roves PM, Geiger PC, Jones TE, Holloszy JO (2007) Exercise-induced mitochondrial biogenesis begins before the increase in muscle PGC-1 α expression. *J Biol Chem* 282(1):194-199

Wulff K, Gatti S, Wettstein JG, Foster RG (2010) Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nat Rev Neurosci* 11(8):589-599

Xi H, Yu Y, Fu Y, Foley J, Halees A, Weng Z (2007) Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. *Genome Res* 17(6):798-806

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 434(7031):338-345

Xu S, Witmer PD, Lumayag S, Kovacs B, Valle D (2007) MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster. *J Biol Chem* 282(34):25053-25066

Yamaguchi S, Mitsui S, Yan LL, Yagita K, Miyake S, Okamura H (2000) Role of DBP in the circadian oscillatory mechanism. *Molec Cell Biol* 20(13):4773-4781

Yamazawa K, Ogata T, Ferguson-Smith AC (2010) Uniparental disomy and human disease: an overview. *Am J Med Genet C Semin Med Genet* 154C(3):329-334

Yan L, Miyake S, Okamura H (2000) Distribution and circadian expression of dbp in SCN and extra-SCN areas in the mouse brain. *J Neurosci Res* 59(2):291-295

Yang JH, Zhang XC, Huang ZP, Zhou H, Huang MB, Zhang S, Chen YQ, Qu LH (2006) snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome. *Nucleic Acids Res* 34(18):5112-5123

Yang SJ, Liang HL, Wong-Riley MT (2006) Activity-dependent transcriptional regulation of nuclear respiratory factor-1 in cultured rat visual cortical neurons. *Neuroscience* 141(3):1181-1192

Yin L, Wu N, Curtin JC, Qatanani M, Szwergold NR, Reid RA, Waitt GM, Parks DJ, Pearce KH, Wisely GB, Lazar MA (2007) Rev-erbalpha, a heme sensor that coordinates metabolic and circadian pathways. *Science* 318(5857):1730-1731

Yoshitane H, Takao T, Satomi Y, Du NH, Okano T, Fukada Y (2009) Roles of CLOCK Phosphorylation in Suppression of E-Box-Dependent Transcription *Mol Cell Biol* 29(13):3675-3686

Young ME, Razeghi P, Taegtmeyer H (2001) Clock genes in the heart: characterization and attenuation with hypertrophy. *Circ Res* 88(11):1142-1150

Young ME, Wilson CR, Razeghi P, Guthrie PH, Taegtmeyer H (2002) Alterations of the circadian clock in the heart by streptozotocin-induced diabetes. *J Mol Cell Cardiol* 34(2):223-231

Zanella S, Barthelemy M, Muscatelli F, Hilaire G (2008) Necdin gene, respiratory disturbances and Prader-Willi syndrome. *Adv Exp Med Biol* 605:159-164

Zemann A, op de Bekke A, Kiefmann M, Brosius J, Schmitz J (2006) Evolution of small nucleolar RNAs in nematodes. *Nucleic Acids Res* 34(9):2676-2685

Zhang J, Fang Z, Jud C, Vansteensel MJ, Kaasik K, Lee CC, Albrecht U, Tamanini F, Meijer JH, Oostra BA, Nelson DL (2008) Fragile X-related proteins regulate mammalian circadian behavioral rhythms. *Am J Hum Genet* 83(1):43-52

Zhao WN, Malinin N, Yang FC, Staknis D, Gekakis N, Maier B, Reischl S, Kramer A, Weitz CJ (2007) CIPC is a mammalian circadian clock protein without invertebrate homologues. *Nat Cell Biol* 9(3):268-275

Zhao X, Yu YT (2008) Targeted pre-mRNA modification for gene silencing and regulation. *Nat Methods* 5(1):95-100

Zhu, W, Samollow PB, Nicholls RD. The Prader-Willi Syndrome (PWS) snoRNAs and those Encoded by Marsupial SNRPB' and SNRPN are paralogous with Divergent Evolution from a Shared Therian ancestor. *Manuscript in preparation*

Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 9(1):133-148